Introduction to Abstract Algebra

Timothy J. Ford

DEPARTMENT OF MATHEMATICS, FLORIDA ATLANTIC UNIVERSITY, BOCA RATON, FL 33431

Email address: ford@fau.edu

URL: http://math.fau.edu/ford

Contents

Preface

This book, or more accurately, these notes, originated in the class notes that I compiled when I taught the two-semester sequence on Abstract Algebra at Florida Atlantic University for the Fall 2019 – Spring 2020 academic year. At my university, the students who take this course are either advanced undergraduates or first year graduate students. Throughout the course, I personally typeset the lecture notes and made them available for my students. Supplemental exercises were added as well. By the end of the course, I had accumulated most of the material in this document. After the course was completed, I organized the somewhat disjoint set of notes into the present form, correcting some cross-references and filling some gaps. This book consists primarily of the notes from my lectures plus material that was added for completeness. It is only fair to mention that a considerable amount of editing has also been performed.

The purpose of this book (or collection of notes) is to provide an introduction to the theory of abstract algebra. The goal is to lay a solid foundation for future study of algebraic topics. It is intended to be accessible to first year graduate students and advanced undergraduate students in mathematics. A typical two-semester sequence on Abstract Algebra at the introductory level would cover most of the material. Chapters two, three, four and five provide a solid introduction to group theory, ring theory, linear algebra and fields. Chapter one, a background chapter, contains much of our conventions concerning notation and terminology as well as a review of the material from set theory and elementary number theory necessary for the rest of the book.

Algebra is one of the fundamental areas of mathematics. Like most of modern mathematics, it is no exaggeration to say that Algebra is very abstract. The many abstract structures and constructions that exist in Algebra can be difficult to grasp upon first encounter. For this reason, it is sometimes helpful to have a "handle" to lend support. In its essence, Algebra is the study of polynomial equations. While not intending to be an oversimplification of the matter, keeping this in mind can be of help to a student trying to make sense of the many abstract notions that arise.

For instance, Number Theory can be considered as that subset of Algebra that is concerned with polynomial equations for which the coefficients involve only natural numbers. Likewise, the origins of Group Theory lie in the study of solutions to polynomial equations in one variable. It was Galois who stressed the importance of looking at the permutations of the set of roots of a polynomial in one indeterminate. This led to what is now called Galois Theory, as well as to the notion of a group acting on a set, hence to what is now called Group Theory.

The set of solutions to a system of polynomials in several variables is called an algebraic variety. Algebraic Geometry arose as the study of algebraic varieties. Linear Algebra is the study of systems of linear equations. Arising out of this study are what we now call vector spaces, and more generally, modules. Matrices turn out to have both practical and theoretical importance in Linear Algebra. Ring Theory can be thought of as the natural abstraction of the addition and multiplication operations possessed by the set of square matrices. Commutative Algebra naturally developed out of the study of properties of rings of functions on algebraic varieties.

CHAPTER 1

Preliminaries and Prerequisites

Chapter 1 is intended to be used as a reference by the subsequent chapters. We assume the reader is familiar with most of the material. This chapter is not intended to be a substitute for an undergraduate textbook on Discrete Mathematics. Conventions, notation and terminology are established. Without undermining the importance of the subject matter, the goal of Chapter 1 is to efficiently and concisely set the table for the rest of the book. Therefore, a practical, or utilitarian approach is taken.

1. Background Material from Set Theory

Sets are the basic building blocks of abstract mathematics. We begin with sets of numbers, sets of letters, sets of sets, or sets of variables. We combine them, operate on them, compare them. Functions, relations and binary operations are themselves defined as sets.

A rigorous definition of a set is not attempted. Rather, we adopt the naive approach that a set is an abstract collection of objects, or elements. It is important to emphasize that the key property or attribute a set is required to possess is that it is possible to distinguish in an unambiguous way those elements that are in the set from those not in the set.

1.1. Sets and operations on sets. A *set* is a collection of objects X with a membership rule such that given any object x it is possible to decide whether x belongs to the set X. If x belongs to X, we say x is an *element* of X and write $x \in X$. Suppose X and Y are sets. If every element of X is also an element of Y, then we say X is a *subset* of Y, or that X is *contained* in Y, and write $X \subseteq Y$. If X and Y are subsets of each other, then we say X and Y are *equal* and write $X = Y$. The set without an element is called the *empty set* and is denoted $\emptyset$. The set of all subsets of X is called the *power set* of X, and is denoted 2^X. Notice that $\emptyset$ and X are both elements of 2^X. The *union* of X and Y, denoted $X \cup Y$, is the set of all elements that are elements of X or Y. The *intersection* of X and Y, denoted $X \cap Y$, is the set of all elements that are elements of X and Y. The *complement* of X with respect to Y, denoted $Y - X$, is the set of all elements of Y that are not elements of X. The *product* of X and Y, denoted $X \times Y$, is the set of all ordered pairs of the form (x, y) where x is an element of X and Y is an element of Y.

Let I be a set and suppose for each $i \in I$ there is a set X_i. Then we say $\{X_i \mid i \in I\}$ is a *family of sets indexed by* I. The *union* of the family is denoted $\bigcup_{i \in I} X_i$ and is defined to be the set of all elements x such that $x \in X_i$ for some $i \in I$. The *intersection* of the family is denoted $\bigcap_{i \in I} X_i$ and is defined to be the set of all elements x such that $x \in X_i$ for all $i \in I$.

The set of *integers* is $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$. The set of *natural numbers* is $\mathbb{N} = \{1, 2, 3, \ldots\}$. The set of nonnegative integers is $\mathbb{Z}_{\geq 0} = \{0, 1, 2, 3, 4, \ldots\}$. The set of *rational numbers* is $\mathbb{Q} = \{n/d \mid n \in \mathbb{Z}, d \in \mathbb{N}\}$ where it is understood that $n/d = x/y$ if $ny = dx$. The set of *real numbers* is denoted $\mathbb{R}$, the set of *complex numbers* is denoted $\mathbb{C}$.

If $n \in \mathbb{N}$ and $\{X_1, \ldots, X_n\}$ is a family of sets indexed by $\{1, 2, \ldots, n\}$, then we sometimes write $X_1 \cup \cdots \cup X_n$ instead of $\bigcup_{i=1}^n X_i$, and $X_1 \cap \cdots \cap X_n$ instead of $\bigcap_{i=1}^n X_i$. The *product* of the family, written $X_1 \times \cdots \times X_n$ or $\prod_{i=1}^n X_i$, is the set $\{(x_1, \ldots, x_n) \mid x_i \in X_i\}$.

1.2. Relations and functions. Let X and Y be nonempty sets. A *relation* between X and Y is a nonempty subset R of the product $X \times Y$. Two relations are equal if they are equal as sets. The *domain* of R is the set of all first coordinates of the pairs in R. The *range* of R is the set of all second coordinates of the pairs in R.

A *function* (or *map*) from X to Y is a relation $f \subseteq X \times Y$ such that for each $x \in X$ there is a unique $y \in Y$ such that $(x, y) \in f$. In this case, we say y is the *image* of x under f, and write $y = f(x)$. The range of a function f is also called the *image* of f. The image of f is denoted $f(X)$, or $\mathrm{im}(f)$. The notation $f : X \to Y$ means f is a function from X to Y. If $T \subseteq Y$, the *preimage* of T under f, denoted $f^{-1}(T)$, is the set of all elements $x \in X$ such that $f(x) \in T$. If $y \in Y$, we usually write $f^{-1}(y)$ instead of $f^{-1}(\{y\})$. If $S \subseteq X$, the *restriction* of f to S is the function $f|_S : S \to Y$ defined by $f|_S(x) = f(x)$ for all $x \in S$. The *identity map* from X to X, $1_X : X \to X$, is defined by $1_X(x) = x$ for all $x \in X$. If $S \subseteq X$, the *inclusion map* from S to X is the restriction of the identity map 1_X to the subset S. If $f : X \to Y$ and $g : Y \to Z$, the *product* or *composition map* is $gf : X \to Z$ defined by $gf(x) = g(f(x))$. If $h : Z \to W$, the reader should verify that $h(gf) = (hg)f$ so the product of functions is associative. We say that $f : X \to Y$ is *one-to-one* (or *injective*) in case $f^{-1}(y)$ is a singleton set for each $y \in f(X)$. We say that $f : X \to Y$ is *onto* or (*surjective*) in case the image of f is equal to Y. If $f : X \to Y$ is one-to-one and onto, then we say that f is a *one-to-one correspondence* (or f is *bijective*). The reader should verify that the identity map 1_X is a one-to-one correspondence. If $S \subseteq X$, the reader should verify that the inclusion map $S \to X$ is one-to-one.

PROPOSITION 1.1.1. *Let $f : X \to Y$.*

(1) f is one-to-one if and only if there exists $g : Y \to X$ such that $gf = 1_X$. In this case g is called a left inverse *of f.*
(2) If f is a one-to-one correspondence, then the function g of Part (1) is unique and satisfies $fg = 1_Y$. In this case g is called the inverse *of f and is denoted f^{-1}.*
(3) If there exists a function $g : Y \to X$ such that $gf = 1_X$ and $fg = 1_Y$, then f is a one-to-one correspondence and g is equal to f^{-1}.

PROOF. (1): View f as a subset of $X \times Y$ and define g as a subset of $Y \times X$. Because f is not onto, our definition of g on $Y - f(X)$ is ad hoc. For this reason, let x_0 be any element of X. Define $g = \{(f(x), x) \mid x \in X\} \cup \{(y, x_0) \mid y \in Y - f(X)\}$. Then g has the desired properties. The rest is Exercise 1.1.8. □

A *commutative diagram* is a finite family of sets $D_V = \{X_1, \ldots, X_v\}$ together with a finite collection of functions $D_E = \{f_1, \ldots, f_e\}$ satisfying the following properties.

(1) Each f in D_E is a function from one set in D_V to another set in D_V.
(2) Given two sets X, Y in D_V and any two paths

$$X = A_0 \xrightarrow{f_{a_1}} A_1 \xrightarrow{f_{a_2}} \cdots \to A_{r-1} \xrightarrow{f_{a_r}} A_r = Y$$

$$X = B_0 \xrightarrow{g_{b_1}} B_1 \xrightarrow{g_{b_2}} \cdots \to B_{s-1} \xrightarrow{g_{b_s}} B_s = Y$$

from X to Y consisting of functions $f_{a_1}, \ldots, f_{a_r}, g_{b_1}, \ldots, g_{b_s}$ in D_E, the composite functions $f_{a_r} \cdots f_{a_1}$ and $g_{b_s} \cdots g_{b_1}$ are equal.

1.3. Binary relations. A *binary relation on X* is a subset of $X \times X$. Suppose $\sim$ is a binary relation on X. If (x,y) is an element of the relation, then we say x is *related* to y and write $x \sim y$. Otherwise we write $x \not\sim y$. If $x \sim x$ for every $x \in X$, then we say $\sim$ is *reflexive*. We say $\sim$ is *symmetric* in case $x \sim y$ whenever $y \sim x$. We say $\sim$ is *antisymmetric* in case $x \sim y$ and $y \sim x$ implies $x = y$. We say $\sim$ is *transitive* if $x \sim z$ whenever $x \sim y$ and $y \sim z$. If $\sim$ is reflexive, symmetric and transitive, then we say $\sim$ is an *equivalence relation* on X. If $\sim$ is an equivalence relation on X, and $x \in X$, then the *equivalence class* containing x is $[x] = \{y \in X \mid x \sim y\}$. By $X/\sim$ we denote the set of all equivalence classes. The function $\eta : X \to X/\sim$ defined by $\eta(x) = [x]$ is called the *natural map*.

PROPOSITION 1.1.2. *Let X be a nonempty set and $\sim$ an equivalence relation on X.*

(1) If $x \in X$, then $[x] \neq \emptyset$.

(2) $\bigcup_{x \in X} [x] = X = \bigcup_{[x] \in X/\sim} [x]$

(3) If $x, y \in X$, then $[x] = [y]$ or $[x] \cap [y] = \emptyset$.

PROOF. Is left to the reader. □

Let X be a nonempty set. A *partition* of X is a family $\mathscr{P}$ of nonempty subsets of X such that $X = \bigcup_{P \in \mathscr{P}} P$ and if $P, Q \in \mathscr{P}$, then either $P = Q$, or $P \cap Q = \emptyset$. If $\sim$ is an equivalence relation on X, then Proposition 1.1.2 shows that $X/\sim$ is a partition of X. Conversely, suppose $\mathscr{P}$ is a partition of X. There is an equivalence relation $\sim$ on X corresponding to $\mathscr{P}$ defined by $x \sim y$ if and only if x and y belong to the same element of $\mathscr{P}$.

PROPOSITION 1.1.3. *Let X be a nonempty set. There is a one-to-one correspondence between the set of all equivalence relations on X and the the set of all partitions of X. The assignment maps an equivalence relation $\sim$ to the partition $X/\sim$.*

PROOF. Is left to the reader. □

Let U be any set, which we assume contains $\mathbb{N}$ as a subset. Define a binary relation on the power set 2^U by the following rule. If X and Y are subsets of U, then we say X and Y are *equivalent* if there exists a one-to-one correspondence $\alpha : X \to Y$. The reader should verify that this is an equivalence relation on 2^U. If X and Y are equivalent sets, then we say X and Y have the same *cardinal number*. Define $I_0 = \emptyset$. For $n \geq 1$ define $I_n = \{1, \ldots, n\}$. If a set X is equivalent to I_n, then we say X has cardinal number n and write $|X| = n$. We say a set X is *finite* if X is equivalent to I_n for some n. Otherwise, we say X is *infinite*.

Let X be a set and $\leq$ a binary relation on X which is reflexive, antisymmetric and transitive. Then we say $\leq$ is a *partial order* on X. We also say X is *partially ordered by* $\leq$. If $x, y \in X$, then we say x and y are *comparable* if $x \leq y$ or $y \leq x$. A *chain* is a partially ordered set with the property that any two elements are comparable. If $S \subseteq X$ is a nonempty subset, then S is partially ordered by the restriction of $\leq$ to $S \times S$. If the restriction of $\leq$ to S is a chain, then we say S is a *chain in X*.

Let X be partially ordered by $\leq$ and suppose S is a nonempty subset of X. Let $a \in S$. We say a is the *least* element of S if $a \leq x$ for all $x \in S$. If it exists, clearly the least element is unique. We say a is a *minimal* element of S in case $x \leq a$ implies $x = a$ for all $x \in S$. We say a is a *maximal* element of S in case $a \leq x$ implies $x = a$ for all $x \in S$. A *well ordered* set is a partially ordered set X such that every nonempty subset S has a least element. The reader should verify that a well ordered set is a chain. An element $u \in X$ is called an *upper bound* for S in case $x \leq u$ for all $x \in S$. An element $l \in X$ is called a *lower bound* for S in case $l \leq x$ for all $x \in S$. An element $U \in X$ is a *supremum*, or *least upper bound* for

S, denoted $U = \sup(S)$, in case U is an upper bound for S and U is a lower bound for the set of all upper bounds for S. The reader should verify that the supremum is unique, if it exists. An element $L \in X$ is an *infimum*, or *greatest lower bound* for S, denoted $L = \inf(S)$, in case L is a lower bound for S and L is an upper bound for the set of all lower bounds for S. The reader should verify that the infimum is unique, if it exists.

Let X be partially ordered by $\leq$. We say that X satisfies the *minimum condition* if every nonempty subset of X contains a minimal element. We say that X satisfies the *maximum condition* if every nonempty subset of X contains a maximal element. We say that X satisfies the *descending chain condition* (DCC) if every chain in X of the form $\{\ldots, x_3 \leq x_2 \leq x_1 \leq x_0\}$ is eventually constant. That is, there is a subscript n such that $x_n = x_i$ for all $i \geq n$. We say that X satisfies the *ascending chain condition* (ACC) if every chain in X of the form $\{x_0 \leq x_1 \leq x_2 \leq x_3, \ldots\}$ is eventually constant.

1.4. Permutations and combinations. Let $n \geq 1$ and $\mathbb{N}_n = \{1, 2, \ldots, n\}$. A bijection $\sigma : \mathbb{N}_n \to \mathbb{N}_n$ is also called a permutation. Let S_n denote the set of all permutations of $\mathbb{N}_n$. In Example 2.1.14 we will call S_n the symmetric group on n letters. If $\sigma \in S_n$, then we can view $\sigma = (x_1, \ldots, x_n)$ as an n-tuple in the product $\prod_{i=1}^{n} \mathbb{N}_n$. The fact that σ is a bijection is equivalent to the statement that the n-tuple $(x_1, \ldots, x_n)$ contains no repeated elements. Therefore,

$$S_n = \left\{ (x_1, \ldots, x_n) \in \prod_{i=1}^{n} \mathbb{N}_n \mid \text{if } i \neq j, \text{ then } x_i \neq x_j \right\}.$$

Because there are n ways to pick x_1, $n-1$ ways to pick x_2, and so forth, a straightforward induction proof shows that the number of elements in S_n is equal to $n!$. If $1 \leq k \leq n$, then a *k-permutation* of $\mathbb{N}_n$ is a one-to-one function $\sigma : \mathbb{N}_k \to \mathbb{N}_n$. The k-permutations of $\mathbb{N}_n$ correspond to k-tuples $(x_1, \ldots, x_k)$ where each $x_i \in \mathbb{N}_n$ and if $i \neq j$, then $x_i \neq x_j$. Again, a straightforward induction proof shows that the number of k-permutations of $\mathbb{N}_n$ is equal to $n(n-1)\cdots(n-k+1) = n!/(n-k)!$.

If X is a finite set with cardinality $|X| = n$, then we say X is an *n-set*. If $S \subseteq X$ and $|S| = k$, then we say S is a *k-subset of X*. The number of k-subsets of an n-set X is denoted $\binom{n}{k}$. The symbol $\binom{n}{k}$ is called the *binomial coefficient* and is pronounced n choose k because it is the number of different ways to choose k objects from a set of n objects.

As we saw above, the number of different k-permutations of $\mathbb{N}_n$ is equal to $n!/(n-k)!$. But a k-permutation of $\mathbb{N}_n$ can be viewed as a two step process. The first step is choosing a k-subset, which can be done in $\binom{n}{k}$ different ways. Then the elements of the k-set are permuted, which can be done in $k!$ ways. Viewing the number of k-permutations of $\mathbb{N}_n$ in these two different ways, we see that $n!/(n-k)!$ is equal to $\binom{n}{k}(k!)$. This leads to Part (3) of the next lemma.

LEMMA 1.1.4. *The following are true.*

(1) If $n < 0$ or $k > n$, then $\binom{n}{k} = 0$.
(2) If $n \geq 0$, then $\binom{n}{0} = \binom{n}{n} = 1$.
(3) If $0 \leq k \leq n$, then $\dbinom{n}{k} = \dfrac{n!}{k!(n-k)!}$.
(4) (Pascal's Identity) If $0 < k < n$, then $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$.

PROOF. Parts (1) and (2) follow straight from the definition of binomial coefficient. Part (3) follows from the paragraph above. Part (4) follows directly from the formula in (3) and is left as an exercise for the reader. □

1.5. Binary operations. Let X be a nonempty set. A *binary operation* on X is a function $X \times X \to X$. If $*$ is a binary operation on X, the image of an ordered pair (x,y) is denoted $x*y$. The binary operation is said to be *associative* if $(x*y)*z = x*(y*z)$ for all $x,y,z \in X$. If e is a special element in X such that $x*e = e*x = x$ for all $x \in X$, then we say e is an *identity element* for $*$. If $x*y = y*x$ for all $x,y \in X$, then we say $*$ is *commutative*. If $(x,y) \mapsto x \cdot y$ is another binary operation on X such that $x \cdot (y*z) = (x \cdot y)*(x \cdot z)$ and $(x*y) \cdot z = (x \cdot z)*(y \cdot z)$ for all $x,y,z \in X$, then we say $\cdot$ *distributes over* $*$.

EXAMPLE 1.1.5. Here are some common examples of binary operations on sets.

(1) Addition of numbers is a binary operation on the set of real numbers $\mathbb{R}$. Addition is associative, commutative, and 0 is the identity element. Multiplication of numbers is a binary operation on the set of real numbers $\mathbb{R}$. Multiplication is associative, commutative, and 1 is the identity element. Multiplication distributes over addition.
(2) Let U be a nonempty set and $X = 2^U$. If A and B are in X, then so are $A \cup B$, $A \cap B$, and $A - B$. In other words, union, intersection, and set difference all define binary operations on X. Union and intersection are both associative and commutative. The distributive laws for union and intersection are in Exercise 1.1.6.
(3) Let X be a nonempty set and $\mathrm{Map}(X)$ the set of all functions mapping X to X. If $f,g \in \mathrm{Map}(X)$, then so is the composite function fg. Composition of functions is a binary operation on $\mathrm{Map}(X)$ which is associative. If $|X| > 1$, then composition of functions in $\mathrm{Map}(X)$ is noncommutative. The identity map 1_X is the identity element.
(4) Let $\mathbb{R}^3 = \{(x_1,x_2,x_3) \mid x_1,x_2,x_3 \in \mathbb{R}\}$ be the set of all ordered 3-tuples over $\mathbb{R}$. The *cross product* of the vector $\mathbf{x} = (x_1,x_2,x_3)$ and the vector $\mathbf{y} = (y_1,y_2,y_3)$ is the vector $\mathbf{x} \times \mathbf{y} = (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1)$. Therefore, cross product is a binary operation on $\mathbb{R}^3$. This binary operation is not associative and not commutative.

1.6. Exercises.

EXERCISE 1.1.6. (Distributive Laws for Intersection and Union) Let $\{X_i \mid i \in I\}$ be a family of sets indexed by I and let Y be any set. Prove:

(1) $Y \cap \left(\bigcup_{i\in I} X_i\right) = \bigcup_{i\in I}(Y \cap X_i)$
(2) $Y \cup \left(\bigcap_{i\in I} X_i\right) = \bigcap_{i\in I}(Y \cup X_i)$

EXERCISE 1.1.7. (DeMorgan's Laws) Let $\{X_i \mid i \in I\}$ be a family of sets indexed by I and suppose U is an arbitrary set. Prove:

(1) $U - \left(\bigcup_{i\in I} X_i\right) = \bigcap_{i\in I}(U - X_i)$
(2) $U - \left(\bigcap_{i\in I} X_i\right) = \bigcup_{i\in I}(U - X_i)$

EXERCISE 1.1.8. Finish the proof of Proposition 1.1.1.

EXERCISE 1.1.9. Let $f : X \to Y$ and $g : Y \to Z$. Prove:

(1) If gf is onto, then g is onto.
(2) If gf is one-to-one, then f is one-to-one.
(3) If f is onto and g is onto, then gf is onto.
(4) If f is one-to-one and g is one-to-one, then gf is one-to-one.

EXERCISE 1.1.10. Recall that the set of natural numbers is $\mathbb{N} = \{1,2,\dots\}$ and if $n \in \mathbb{N}$, then $\mathbb{N}_n = \{1,2,\dots,n\}$. Prove:

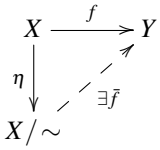
(1) If $f : \mathbb{N}_n \to \mathbb{N}_n$ is one-to-one, then f is onto.
(2) If $f : \mathbb{N}_n \to \mathbb{N}_n$ is onto, then f is one-to-one.

EXERCISE 1.1.11. (The Pigeonhole Principle) Let $f : \mathbb{N}_m \to \mathbb{N}_n$. Prove:

(1) If $m > n$, then f is not one-to-one.
(2) If $m < n$, then f is not onto.

EXERCISE 1.1.12. Let X and Y be finite sets. Show that $|X \times Y| = |X||Y|$.

EXERCISE 1.1.13. (Universal Mapping Property) Let $f : X \longrightarrow Y$ be a function. Let $\sim$ be an equivalence relation on X, and $\eta : X \to X/\sim$ the natural map. Show that if f has the property that $a \sim b$ implies $f(a) = f(b)$ for all $a, b \in X$, then there exists a function $\bar{f} : X/\sim \to Y$ such that $f = \bar{f}\eta$. Hence the diagram



commutes. This shows that if f is constant on equivalence classes, then f factors through the natural map η.

EXERCISE 1.1.14. Let $f : X \longrightarrow Y$ be a function. Define a relation $\approx$ on X by the rule: $x \approx y$ if and only if $f(x) = f(y)$. Prove:

(1) $\approx$ is an equivalence relation on X.
(2) There exists a function $\bar{f} : X/\approx \to Y$ such that f factors through the natural map $\eta : X \to X/\approx$. That is, $f = \bar{f}\eta$.
(3) $\bar{f}$ is one-to-one.
(4) $\bar{f}$ is a one-to-one correspondence if and only if f is onto.

EXERCISE 1.1.15. Let X be an infinite set. Prove that X contains a subset that is equivalent to $\mathbb{N}$.

EXERCISE 1.1.16. Let X be a set. Prove that X is infinite if and only if there exists a one-to-one function $f : X \to X$ which is not onto.

EXERCISE 1.1.17. If $x \in \mathbb{R}$, the *floor* of x, written $\lfloor x \rfloor$, is the maximum of the set $\{k \in \mathbb{Z} \mid k \leq x\}$. The *ceiling* of x, written $\lceil x \rceil$, is the minimum of the set $\{k \in \mathbb{Z} \mid k \geq x\}$. Let $f : \mathbb{N}_m \to \mathbb{N}_n$. Prove:

(1) There exists $a \in \mathbb{N}_n$ such that the cardinality of the set $f^{-1}(a)$ is greater than or equal to $\lceil m/n \rceil$.
(2) There exists $b \in \mathbb{N}_n$ such that the cardinality of the set $f^{-1}(b)$ is less than or equal to $\lfloor m/n \rfloor$.

EXERCISE 1.1.18. Prove the Binomial Theorem:

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

where x and y are indeterminates and $n \geq 0$.

EXERCISE 1.1.19. Let X be a finite set. Use the Binomial Theorem to prove that $|2^X| = 2^{|X|}$.

2. Background Material from Number Theory

The basic results from Elementary Number Theory that will be required later in the text are included here. The set of integers is $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$. We assume the reader is familiar with its partial ordering, and the binary operations of addition and multiplication. No attempt is made to construct the integers from from first principles. The set of natural numbers is $\mathbb{N} = \{1, 2, 3, \dots\}$. The Well Ordering Principle is assumed as an axiom.

AXIOM 1.2.1. *(The Well Ordering Principle) If S is a nonempty subset of $\mathbb{Z}$ and S has a lower bound, then S contains a least element.*

PROPOSITION 1.2.2. *(Mathematical Induction) Let S be a subset of $\mathbb{N}$ such that $1 \in S$. Assume S satisfies one of the following.*

(1) For each $n \in \mathbb{N}$, if $n \in S$, then $n+1 \in S$.
(2) For each $n \in \mathbb{N}$, if $\{1, \dots, n\} \subseteq S$, then $n+1 \in S$.

Then $S = \mathbb{N}$.

PROOF. Assume $S \subseteq \mathbb{N}$, $1 \in S$, and S satisfies (1) or (2). Let $C = \mathbb{N} - S$. For contradiction's sake assume $C \neq \emptyset$. By Axiom 1.2.1, C has a least element, say ℓ. Since $1 \in S$, we know $\ell > 1$. Therefore, $\ell - 1 \in S$ and $\ell \notin S$, which contradicts (1). Since ℓ is the least element of C, $\{1, \dots, \ell - 1\} \subseteq S$ and $\ell \notin S$, which contradicts (2). We conclude that $C = \emptyset$, hence $S = \mathbb{N}$. □

PROPOSITION 1.2.3. *(The Division Algorithm) If $a, b \in \mathbb{Z}$ and $a \neq 0$, then there exist unique integers $q, r \in \mathbb{Z}$ such that $0 \leq r < |a|$ and $b = aq + r$.*

PROOF. First we prove the existence claim. Let $S = \{b - ax \mid x \in \mathbb{Z} \text{ and } b - ax \geq 0\}$. If $x > |b|$, then it follows that $b + |a|x \geq 0$. Therefore, either $b + ax$ or $b - ax$ is in S. By Axiom 1.2.1, S has a least element, say $r = b - aq$, for some $q \in \mathbb{Z}$. For contradiction's sake, assume $r \geq |a|$. Then $0 \leq r - |a| = b - aq - |a| = b - a(q \pm 1)$. This implies $r - |a| \in S$, contradicting the minimal choice of r.

To prove the uniqueness claim, suppose $b = aq + r = aq_1 + r_1$ and $0 \leq r \leq r_1 < |a|$. Then $|r_1 - r| = |a||q - q_1|$. Since $0 \leq r_1 - r < |a|$, this implies $q - q_1 = 0$. Hence $r_1 - r = 0$. □

Let $a, b \in \mathbb{Z}$. We say *a divides b*, and write $a \mid b$, in case there exists $q \in \mathbb{Z}$ such that $b = aq$. In this case, a is called a *divisor* of b, and b is called a *multiple* of a.

PROPOSITION 1.2.4. *Let $\{a_1, \dots, a_n\}$ be a set of integers and assume at least one of the a_i is nonzero. There exists a unique positive integer d such that*

(1) $d \mid a_i$ for all $1 \leq i \leq n$, and
(2) if $e \mid a_i$ for all $1 \leq i \leq n$, then $e \mid d$.

We call d the greatest common divisor *of the set, and write $d = \gcd(a_1, \dots, a_n)$.*

PROOF. Let S be the set of all positive linear combinations of the a_i

$$S = \{x_1 a_1 + \cdots + x_n a_n \mid x_1, \dots, x_n \in \mathbb{Z}, x_1 a_1 + \cdots + x_n a_n > 0\}.$$

The reader should verify that $S \neq \emptyset$. By Axiom 1.2.1, there exists a least element of S which we can write as $d = k_1 a_1 + \cdots + k_n a_n$ for some integers $k_1, \dots, k_n$. Fix one i and apply the Division Algorithm to write $a_i = dq + r$ where $0 \leq r < d$. Solve $a_i = (k_1 a_1 + \cdots + k_n a_n)q + r$ for r to see that

$$r = a_i - (k_1 a_1 + \cdots + k_n a_n)q$$

is a linear combination of $a_1, \ldots, a_n$. Because $r < d$, we conclude that r is not in S. Therefore $r = 0$. This proves Part (1). The reader should verify Part (2) and the claim that d is unique. □

We say the set of integers $\{a_1, \ldots, a_n\}$ is *relatively prime* in case $\gcd(a_1, \ldots, a_n) = 1$. An integer $\pi \in \mathbb{Z}$ is called a *prime* in case $\pi > 1$ and the only divisors of π are $-\pi, -1, 1, \pi$.

LEMMA 1.2.5. *Let a, b and c be integers. Assume $a \neq 0$ or $b \neq 0$.*

(1) (Bézout's Identity) If $d = \gcd(a,b)$, then there exist integers u and v such that $d = au + bv$.
(2) (Euclid's Lemma) If $\gcd(a,b) = 1$ and $a \mid bc$, then $a \mid c$.
(3) If there exist integers u and v such that $1 = au + bv$, then $\gcd(a,b) = 1$.

PROOF. (1): This is immediate from the proof of Proposition 1.2.4.
(2): Assume $\gcd(a,b) = 1$. By Part (1) there exist integers u and v such that $1 = au + bv$. Then $c = acu + bcv$. Since a divides the right hand side, a divides c.
(3): This is immediate from the proof of Proposition 1.2.4. □

LEMMA 1.2.6. *Let π be a prime number. Let a and $a_1, \ldots, a_n$ be integers.*

(1) If $\pi \mid a$, then $\gcd(\pi, a) = \pi$, otherwise $\gcd(\pi, a) = 1$.
(2) If $\pi \mid a_1 a_2 \cdots a_n$, then $\pi \mid a_i$ for some i.

PROOF. (1): The proof is an exercise for the reader.
(2): For sake of contradiction, assume the statement is false. Let π and $a_1, \ldots, a_n$ be a counterexample such that n is minimal. Then π divides the product $a_1 \cdots a_n$ and by (1) $\gcd(\pi, a_i) = 1$ for each i. Again by (1), $n > 1$. By Lemma 1.2.5 applied to $a_1(a_2 \cdots a_n)$, $\pi \mid a_2 \cdots a_n$. By the minimal choice of n, π divides one of $a_2, \ldots, a_n$. This is a contradiction. □

PROPOSITION 1.2.7. *(The Fundamental Theorem of Arithmetic) Let n be a positive integer which is greater than 1. There exist unique positive integers k, $e_1, \ldots, e_k$ and unique prime numbers $p_1, \ldots, p_k$ such that $n = p_1^{e_1} \cdots p_k^{e_k}$.*

PROOF. First we prove the existence claim. If n is a prime, then set $k = 1$, $p_1 = n$, $e_1 = 1$, and we are done. In particular, the result is true for $n = 2$. The proof is by induction on n. Assume that every number in the set $\{2, 3, \ldots, n-1\}$ has a representation as a product of primes. Assume $n = xy$ is composite and that $2 \leq x \leq y \leq n-1$. By the induction hypothesis, both x and y have representations as products of primes. Then $n = xy$ also has such a representation. By Proposition 1.2.2, we are done.

For the uniqueness claim, assume

$$n = p_1^{e_1} \cdots p_k^{e_k} = q_1^{f_1} \cdots q_\ell^{f_\ell} \tag{2.1}$$

are two representations of n as products of primes. Let $M = \sum_{i=1}^{k} e_i$ and $N = \sum_{i=1}^{\ell} f_i$. Without loss of generality, assume $M \leq N$. The proof is by induction on M. If $M = 1$, then $n = p_1$ is prime. This implies $\ell = 1 = f_1$ and $q_1 = p_1$. Assume inductively that $M > 1$ and that the uniqueness claim is true for any product involving $M - 1$ factors. Using Lemma 1.2.6 we see that p_1 divides one of the q_i. Since q_i is prime, this implies p_1 is equal to q_i. Canceling p_1 and q_i from both sides of Eq.(2.1) results in a product of primes with $M - 1$ factors. By the induction hypothesis, we conclude that $k = \ell$ and the sets $\{p_1^{e_1}, \ldots, p_k^{e_k}\}$ and $\{q_1^{f_1}, \ldots, q_k^{f_k}\}$ are equal. □

DEFINITION 1.2.8. Let m be a positive integer. Define a binary relation on $\mathbb{Z}$ by the following rule. Given $x,y \in \mathbb{Z}$, we say *x is congruent to y modulo m*, and write $x \equiv y \pmod{m}$, in case $m \mid (x-y)$. By Proposition 1.2.9 this defines an equivalence relation on $\mathbb{Z}$. The set of all equivalence classes of integers modulo m is denoted $\mathbb{Z}/(m)$.

PROPOSITION 1.2.9. *Let m be a positive integer.*

(1) Congruence modulo m is an equivalence relation on $\mathbb{Z}$.

(2) $\{0,1,\ldots,m-1\}$ is a full set of representatives for the equivalence classes. In other words, every integer is congruent to one of $0,1,\ldots,m-1$ and no two distinct elements of $\{0,1,\ldots,m-1\}$ are congruent to each other.

(3) If $u \equiv v \pmod{m}$ and $x \equiv y \pmod{m}$, then $u+x \equiv v+y \pmod{m}$ and $ux \equiv vy \pmod{m}$.

(4) If $\gcd(a,m) = 1$ and $ax \equiv ay \pmod{m}$, then $x \equiv y \pmod{m}$.

PROOF. (1): Since $m \mid 0$, $x \equiv x \pmod{m}$ for every $x \in \mathbb{Z}$. If $x-y = mq$, then $y-x = m(-q)$. Therefore, $x \equiv y \pmod{m}$ implies $y \equiv x \pmod{m}$. If $x-y = mq$ and $y-z = mr$, then adding yields $x-z = m(q+r)$. Therefore, $x \equiv y \pmod{m}$ and $y \equiv z \pmod{m}$ implies $x \equiv z \pmod{m}$.

(2): By Proposition 1.2.3, if $x \in \mathbb{Z}$, then there exist unique integers q and r such that $x = mq+r$ and $0 \le r < m$. This implies $x \equiv r \pmod{m}$, and $\mathbb{Z}/(m) \subseteq \{0,1,\ldots,m-1\}$. Equality of the two sets follows from the uniqueness of q and r.

(3): Write $u-v = mq$ and $x-y = mr$ for integers q,r. Adding, we get $u-v+x-y = (u+x)-(v+y) = m(q+r)$, hence $u+x \equiv v+y \pmod{m}$. Multiplying the first equation by x and the second by v we have $ux-vx = mxq$ and $xv-yv = mvr$. Adding, we get $ux-vx+xv-yv = ux-yv = m(xq+vr)$, hence $ux \equiv vy \pmod{m}$.

(4): By Lemma 1.2.5 we write $1 = au+mv$ for integers u,v. We are given that $a(x-y) = mq$ for some integer q. Multiply by u to get $au(x-y) = muq$. Substitute $au = 1-mv$ and rearrange to get $x-y = mv(x-y)+muq$. Hence $x \equiv y \pmod{m}$. □

If $a,b \in \mathbb{Z}-\{0\}$, then $|ab| \in S$ is a common multiple of both a and b. Therefore, the set $S = \{x \in \mathbb{Z} \mid a \mid x,\ b \mid x \text{ and } x > 0\}$ is nonempty. By Axiom 1.2.1, S has a least element, which is called the *least common multiple* of a and b, and is denoted $\operatorname{lcm}(a,b)$.

PROPOSITION 1.2.10. *Suppose $a > 0$ and $b > 0$. Then the following are true.*

(1) If $c \in \mathbb{Z}$ and $a \mid c$ and $b \mid c$, then $\operatorname{lcm}(a,b) \mid c$.

(2) $\gcd(a,b)\operatorname{lcm}(a,b) = ab$.

PROOF. (1): Let $\operatorname{lcm}(a,b) = L$. By Proposition 1.2.3, $c = Lq+r$ where $0 \le r < L$. Since $a \mid c$ and $a \mid L$, we see that a divides $r = c-Lq$. Likewise, $b \mid c$ and $b \mid L$ implies that b divides r. So r is a common multiple of a and b and $r < L$. By the definition of L, we conclude that $r = 0$.

(2): Write $d = \gcd(a,b)$. Then $(ab)/d = a(b/d) = (a/d)b$ is a common multiple of a and b. By (1), $L \mid (ab)/d$, or equivalently, $dL \mid ab$. By Lemma 1.2.5, $d = ax+by$ for some integers x,y. Multiply by L to get $dL = aLx+bLy$. Since L is a common multiple of a and b we see that ab divides $aLx+bLy = dL$. We have shown that $dL \mid ab$ and $ab \mid dL$. Both numbers are positive, so we have equality. □

THEOREM 1.2.11. *(Chinese Remainder Theorem) Let m and n be relatively prime positive integers. Then the function*

$$\mathbb{Z}/mn \xrightarrow{\psi} \mathbb{Z}/m \times \mathbb{Z}/n$$

defined by $\psi([x]) = ([x],[x])$ is a one-to-one correspondence.

PROOF. We know that ψ is well defined, by Exercise 1.2.19. By Exercise 1.1.12 and Proposition 1.2.9, $|\mathbb{Z}/m \times \mathbb{Z}/n| = |\mathbb{Z}/mn| = mn$. By Exercise 1.1.10, it is enough to show ψ is one-to-one. Suppose $\psi([x]) = \psi([y])$. Then $x \equiv y \pmod{m}$ and $x \cong y \pmod{n}$, which implies $x - y$ is a common multiple of m and n. By Proposition 1.2.10, $x - y$ is divisible by $\operatorname{lcm}(m,n)$. But $\operatorname{lcm}(m,n) = mn$ since $\gcd(a,b) = 1$. This implies $x \equiv y \pmod{mn}$, and we have shown that ψ is one-to-one. □

Let $n \geq 1$. By Exercise 1.2.20, if $x \equiv y \pmod{n}$, then $\gcd(x,n) = \gcd(y,n)$. This says the function $\mathbb{Z} \to \mathbb{Z}$ defined by $x \mapsto \gcd(x,n)$ is constant on congruence classes. The set $U_n = \{[k] \in \mathbb{Z}/n \mid \gcd(k,n) = 1\}$ is called the set of *units modulo n*. The Euler ϕ-function is defined to be the number of units modulo n. That is, $\phi(n) = |U_n|$. In the terminology of Definition 2.1.1, Lemma 1.2.12 shows that U_n is an abelian group of order $\phi(n)$.

LEMMA 1.2.12. *Let $n \geq 1$.*

(1) If $[a] \in U_n$, then there exists $[b] \in U_n$ such that $[a][b] = [1]$.
(2) If $a, b \in \mathbb{Z}$ and $ab \equiv 1 \pmod{n}$, then $[a] \in U_n$ and $[b] \in U_n$.

PROOF. (1): If $[a] \in U_n$, then $\gcd(a,n) = 1$. By Lemma 1.2.5, there exist integers b, c such that $ab + nv = 1$. Therefore, $ab \equiv 1 \pmod{n}$.

(2): If $ab \equiv 1 \pmod{n}$, then $ab = nq + 1$ for some integer q. By Lemma 1.2.5, $\gcd(a,n) = 1$ and $\gcd(b,n) = 1$. □

PROPOSITION 1.2.13. *If p is a prime and $k \geq 1$, then $\phi(p^k) = p^k - p^{k-1} = p^k(1 - 1/p)$.*

PROOF. The multiples of p in the set $\{1, 2, \ldots, p^k\}$ are $p, 2p, \ldots, p^{k-1}p$. Since there are p^{k-1} multiples of p, there are $p^k - p^{k-1}$ numbers that are relatively prime to p. □

PROPOSITION 1.2.14. *Let m and n be relatively prime positive integers. Then $\phi(mn) = \phi(m)\phi(n)$.*

PROOF. By Theorem 1.2.11, the function $\psi : \mathbb{Z}/mn \to \mathbb{Z}/m \times \mathbb{Z}/n$ defined by $\psi([x]) = ([x],[x])$ is a one-to-one correspondence. We show that the restriction of ψ to U_{mn} induces a one-to-one correspondence $\rho : U_{mn} \to U_m \times U_n$.

If $\gcd(x, mn) = 1$, then by Lemma 1.2.5 there exist integers u, v such that $1 = xu + mnv$, hence $\gcd(x,m) = 1$ and $\gcd(x,n) = 1$. This proves that ρ is well defined. Since ψ is one-to-one, so is ρ. To finish the proof we show that ρ is onto. Let $([a],[b]) \in U_m \times U_n$. By Lemma 1.2.12 there exists $([x],[y]) \in U_m \times U_n$ such that $ax \equiv 1 \pmod{m}$ and $by \equiv 1 \pmod{n}$. Since ψ is onto, there exists $[k] \in \mathbb{Z}/mn$ such that $k \equiv a \pmod{m}$ and $k \equiv b \pmod{n}$. Likewise, there exists $[\ell] \in \mathbb{Z}/mn$ such that $\ell \equiv x \pmod{m}$ and $\ell \equiv y \pmod{n}$. By Proposition 1.2.9, $k\ell \equiv ax \equiv 1 \pmod{m}$ and $k\ell \equiv by \equiv 1 \pmod{n}$. Since ψ is one-to-one, $k\ell \equiv 1 \pmod{mn}$. By Lemma 1.2.12 this implies $[k] \in U_{mn}$, which proves ρ is onto. □

DEFINITION 1.2.15. Let $n \geq 1$ be an integer. The notation $\sum_{d|n}$ or $\prod_{d|n}$ denotes the sum or product over the set of all positive numbers d such that $d \mid n$. An integer n is said to be *square free* if for every prime p, n is not a multiple of p^2. The Möbius function is defined by

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n \text{ is not square free,} \\ (-1)^r & \text{if } n \text{ factors into } r \text{ distinct primes.} \end{cases}$$

THEOREM 1.2.16. *(Möbius Inversion Formula) Let f be a function defined on $\mathbb{N}$ and define another function on $\mathbb{N}$ by*

$$F(n) = \sum_{d|n} f(d).$$

Then

$$f(n) = \sum_{d|n} \mu(d) F\left(\frac{n}{d}\right).$$

PROOF. The proof can be found in any elementary number theory book, and is left to the reader. □

2.1. Exercises.

EXERCISE 1.2.17. Let a and b be integers that are not both zero and let d be the greatest common divisor of a and b. Consider the linear diophantine equation: $d = ax+by$. Bézout's Identity says that there exist integers u and v such that $d = au+bv$.

(1) Show that the matrix $\begin{pmatrix} u & v \\ u-b/d & v+a/d \end{pmatrix}$ is invertible. Find its inverse.
(2) If c is an integer, show that the linear diophantine equation $c = ax+by$ has a solution if and only if $d \mid c$.
(3) Assume $d \mid c$. Prove that the general solution to the linear diophantine equation $c = ax+by$ is $x = x_0 - tb/d$, $y = y_0 + ta/d$, where $t \in \mathbb{Z}$ and (x_0, y_0) is any particular solution.

EXERCISE 1.2.18. This exercise is based on Problem 1.3 of Adrian Wadsworth's book [**15**]. Let a and b be relatively prime positive integers and consider the set

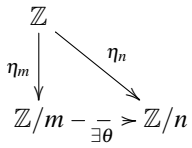
$$L = \{ax+by \mid x \text{ and } y \text{ are nonnegative integers}\}.$$

The problem is to find the integer ℓ satisfying these two properties: (1) $\ell - 1 \notin L$ and (2) if n is an integer and $n \geq \ell$, then $n \in L$.

You are encouraged to solve this interesting problem yourself. Alternatively, you may follow the six steps below which outline a solution.

(1) Prove that if $a = 1$ or $b = 1$, then L contains the set of all nonnegative integers.
(2) Prove that the integers a, b, ab, $(a-1)(b-1)$ are in L.
(3) Prove that $ab-a-b = (a-1)(b-1)-1$ is not in L. Hint: Show that the line $ab-a-b = ax+by$ contains the two lattice points $(-1, a-1)$ and $(b-1, -a)$.
(4) Prove that if $n \geq ab$, then n is in L.
(5) Assume $a > 1$, $b > 1$, and let n be an integer satisfying $ab-a-b < n < ab$. Prove that n is in L. Hints: For sake of contradiction assume $ab-a-b < n < ab$ and n is not in L. Show that there exists an ordered pair (x_1, y_1) such that $n = ax_1+by_1$, (x_1, y_1) is in Quadrant IV and (x_1-b, y_1+a) is in Quadrant II. Show that (x_1, y_1) is not in the parallelogram with vertices $(b,0)$, $(0,a)$, $(-1, a-1)$, $(b-1, -1)$. Show that this is impossible.
(6) Let $\ell = (a-1)(b-1)$. Prove that $\ell - 1 \notin L$ and if $\ell \leq n$, then n is in L.

EXERCISE 1.2.19. Let $m, n \in \mathbb{N}$. Consider the diagram



where η_m and η_n are the natural maps. Show that there exists a function θ making the diagram commute if and only if n divides m.

EXERCISE 1.2.20. Let $n \geq 1$. Show that the function $\mathbb{Z} \to \mathbb{Z}$ defined by $x \mapsto \gcd(x,n)$ is constant on congruence classes. In other words, show that $x \equiv y \pmod{n}$ implies $\gcd(x,n) = \gcd(y,n)$.

EXERCISE 1.2.21. Let p be a prime.

(1) If $1 \leq k \leq p-1$, show that p divides $\binom{p}{k}$.
(2) Show that $(a+b)^p \equiv a^p + b^p \pmod{p}$ for any integers a and b. (Hint: Exercise 1.1.18.)
(3) Use (2) and Proposition 1.2.2 to prove that $(a+b)^{p^n} \equiv a^{p^n} + b^{p^n} \pmod{p}$ for any integers a and b and for all $n \geq 0$.

See Exercise 3.6.31 for a generalization of this result.

EXERCISE 1.2.22. Show that the Möbius function μ is multiplicative in the sense that if $\gcd(m,n) = 1$, then $\mu(mn) = \mu(m)\mu(n)$.

EXERCISE 1.2.23. Let $n \geq 0$ and $X = \prod_{i=1}^{n} \mathbb{Z}_{\geq 0} = \{(x_1, \ldots, x_n) \mid x_i \in \mathbb{Z}_{\geq 0}\}$, where $\mathbb{Z}_{\geq 0} = \{x \in \mathbb{Z} \mid x \geq 0\}$ is the set of nonnegative integers. The *lexicographical ordering* (also called alphabetical or dictionary ordering) on X is defined recursively on n. For $n = 1$, the usual ordering on $\mathbb{Z}$ is applied. If $n > 1$, then $(v_1, v_2, \ldots, v_n) < (w_1, w_2, \ldots, w_n)$ if and only if: $(v_1, v_2, \ldots, v_{n-1}) < (w_1, w_2, \ldots, w_{n-1})$ or $(v_1, v_2, \ldots, v_{n-1}) = (w_1, w_2, \ldots, w_{n-1})$ and $v_n < w_n$. If $\alpha, \beta \in X$, then we write $\alpha \leq \beta$ in case $\alpha < \beta$ or $\alpha = \beta$.

(1) Show that $\leq$ is a partial order on X. Show that X is a chain.
(2) If $\alpha \in X$, the *segment of X determined by α*, written $(-\infty, \alpha)$, is $\{x \in X \mid x < \alpha\}$. For which $\alpha \in X$ is
    (a) $(-\infty, \alpha) = \emptyset$?
    (b) $(-\infty, \alpha)$ finite?
    (c) $(-\infty, \alpha)$ infinite?
(3) Show that X with the lexicographical ordering $\leq$ is a well ordered set. That is, show that if $S \subseteq X$ and $S \neq \emptyset$, then S has a least element.

EXERCISE 1.2.24. Let $X = \{x_0, x_1, \ldots, x_{n-1}\}$ be a finite set and $\mathbb{Z}_{\geq 0}$ the set of nonnegative integers. If $U \subseteq X$, the so-called *indicator function* on U, denoted $\chi_U : U \to \{0,1\}$, is defined by

$$\chi_U(x) = \begin{cases} 1 & \text{if } x \in U \\ 0 & \text{if } x \notin U. \end{cases}$$

Define $f : 2^X \to \mathbb{Z}_{\geq 0}$ by $f(U) = \sum_{i=0}^{n-1} \chi_U(x_i) 2^i$. Prove:

(1) f is a one-to-one correspondence between 2^X and $\{0, 1, \ldots, 2^{n-1}\}$.
(2) $|2^X| = 2^{|X|}$.
(3) The ordering on 2^X induced by the function f makes 2^X into a well ordered set.

3. The Well Ordering Principle and Some of Its Equivalents

Most readers will prefer to make a quick scan of this section on first reading. It is included for completeness' sake as well as a tribute to the influence of [**9**, Chapter 0, Theorem 25] on the author's fondness for the subject. In this book, the only direct application of Zorn's Lemma, Proposition 1.3.3, is in the proof that a commutative ring contains a maximal ideal (see Proposition 3.2.26). As a historical note, Zorn's Lemma, which is equivalent

to the Well Ordering Principle, has been called Zorn's Lemma since M. Zorn first used it to prove that a commutative ring contains a maximal ideal [**17**]. The Axiom of Choice, Proposition 1.3.5, guarantees that a product of nonempty sets is nonempty, but throughout this book we limit our applications to products of at most a countably infinite number of sets. Aside from the application to show the existence of maximal ideals, the other applications of the Well Ordering Principle or one of its equivalents appear in several exercises that are inserted to challenge the reader.

Although we do not prove it here, the Well Ordering Principle, the Principle of Transfinite Induction, Zorn's Lemma, and the Axiom of Choice are logically equivalent to each other.

AXIOM 1.3.1. *(The Well Ordering Principle) If X is a nonempty set, then there exists a partial order $\leq$ on X such that X is a well ordered set. That is, every nonempty subset of X has a least element.*

Let X be a set and $\leq$ a partial order on X. If $x, y \in X$, then we write $x < y$ in case $x \leq y$ and $x \neq y$. Suppose $C \subseteq X$ is a chain in X and $\alpha \in C$. The *segment of C determined by α*, written $(-\infty, \alpha)$, is the set of all elements $x \in C$ such that $x < \alpha$. A subset $W \subseteq C$ is called an *inductive subset* of C provided that for any $\alpha \in C$, if $(-\infty, \alpha) \subseteq W$, then $\alpha \in W$.

PROPOSITION 1.3.2. *(The Transfinite Induction Principle) Suppose X is a well ordered set and W is an inductive subset of X. Then $W = X$.*

PROOF. Suppose $X - W$ is nonempty. Let α be the least element of $X - W$. Then W contains the segment $(-\infty, \alpha)$. Since W is inductive, it follows that $\alpha \in W$, which is a contradiction. □

PROPOSITION 1.3.3. *(Zorn's Lemma) Let X be a partially ordered set. If every chain in X has an upper bound, then X contains a maximal element.*

PROOF. By Axiom 1.3.1, there exists a well ordered set W and a one-to-one correspondence $\omega : W \to X$. Using Proposition 1.3.2, define a sequence $\{C(w) \mid w \in W\}$ of well ordered subsets of X. If w_0 is the least element of W, define $C(w_0) = \{\omega(w_0)\}$. Inductively assume $\alpha \in W - \{w_0\}$ and that for all $w < \alpha$, $C(w)$ is defined and the following are satisfied

(1) if $w_0 \leq w_1 \leq w_2 < \alpha$, then $C(w_1) \subseteq C(w_2)$,
(2) $C(w)$ is a well ordered chain in X, and
(3) $C(w) \subseteq \{\omega(i) \mid w_0 \leq i \leq w\}$.

Let $x = \omega(\alpha)$ and

$$F = \bigcup_{w<\alpha} C(w).$$

The reader should verify that F is a well ordered chain in X and $F \subseteq \{\omega(i) \mid w_0 \leq i < \alpha\}$. Define $C(\alpha)$ by the rule

$$C(\alpha) = \begin{cases} F \cup \{x\} & \text{if } x \text{ is an upper bound for } F \\ F & \text{otherwise.} \end{cases}$$

The reader should verify that $C(\alpha)$ satisfies

(4) if $w_0 \leq w_1 \leq w_2 \leq \alpha$, then $C(w_1) \subseteq C(w_2)$,
(5) $C(\alpha)$ is a well ordered chain in X, and
(6) $C(\alpha) \subseteq \{\omega(i) \mid w_0 \leq i \leq \alpha\}$.

By Proposition 1.3.2, the sequence $\{C(w) \mid w \in W\}$ is defined and the properties (4), (5) and (6) are satisfied for all $\alpha \in W$. Now set

$$G = \bigcup_{w<\alpha} C(w).$$

The reader should verify that G is a well ordered chain in X. By hypothesis, G has an upper bound, say u. We show that u is a maximal element of X. For contradiction's sake, assume X has no maximal element. Then we can choose the upper bound u to be an element of $X - G$. For some $w_1 \in W$ we have $u = \omega(w_1)$. For all $w < w_1$, u is an upper bound for $C(w)$. By the definition of $C(w_1)$, we have $u \in C(w_1)$. This is a contradiction, because $C(w_1) \subseteq G$. □

DEFINITION 1.3.4. Let I be a set and $\{X_i \mid i \in I\}$ a family of sets indexed by I. The *product* is

$$\prod_{i\in I} X_i = \Big\{ f : I \to \bigcup X_i \mid f(i) \in X_i \Big\}.$$

An element f of the product is called a choice function, because f chooses one element from each member of the family of sets.

PROPOSITION 1.3.5. *(The Axiom of Choice) Let I be a set and $\{X_i \mid i \in I\}$ a family of nonempty sets indexed by I. Then the product $\prod_{i\in I} X_i$ is nonempty. That is, there exists a function f on I such that $f(i) \in X_i$ for each $i \in I$.*

PROOF. By Axiom 1.3.1, we can assume $\bigcup_{i\in I} X_i$ is well ordered. We can view X_i as a subset of $\bigcup_{i\in I} X_i$. For each $i \in I$, let x_i be the least element of X_i. The set of ordered pairs (i, x_i) defines the choice function. □

3.1. Exercises.

EXERCISE 1.3.6. Let I be a set and $\{X_i \mid i \in I\}$ a family of nonempty sets indexed by I. For each $k \in I$ define $\pi_k : \prod_{i\in I} X_i \to X_k$ by the rule $\pi_k(f) = f(k)$. We call π_k the *projection onto coordinate k*. Show that π_k is onto.

EXERCISE 1.3.7. Let X be a set that is partially ordered by $\leq$.

(1) Prove that X satisfies the descending chain condition (DCC) if and only if X satisfies the minimum condition.
(2) Prove that X satisfies the ascending chain condition (ACC) if and only if X satisfies the maximum condition.

EXERCISE 1.3.8. Use the Axiom of Choice to prove: A function $f : X \to Y$ is onto if and only if there exists a function $g : Y \to X$ such that $fg = 1_Y$. In this case g is called a *right inverse* of f.

4. Background Material from Calculus

As in Section 1.1.1, the set of real numbers is denoted $\mathbb{R}$.

THEOREM 1.4.1. *If a is a positive real number, then there exists a real number x such that $x^2 = a$. In other words, a positive real number has a square root.*

PROOF. See, for instance, [**13**, Theorem 7.8, p. 124]. □

THEOREM 1.4.2. *If n is a positive odd integer and $a_0, a_1, \ldots, a_{n-1}$ are real numbers, then there exists a real number x such that $x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0$. In other words, a polynomial over $\mathbb{R}$ of odd degree has a root.*

PROOF. See, for instance, [**13**, Theorem 7.9, p. 125]. □

For properties of the complex numbers, the reader is referred, for example, to [**13**, Chapter 25]. The set of complex numbers, denoted $\mathbb{C}$, is identified with $\mathbb{R}^2$ and is a two-dimensional real vector space. A complex number is an ordered pair (a,b). A basis for $\mathbb{C}$ is $(1,0)$, also denoted 1, and $(0,1)$, also denoted i. In terms of this basis, the complex number (a,b) has representation $a+bi$. Addition of complex numbers is coordinate-wise: $(a+bi)+(c+di)=(a+c)+(b+d)i$. The additive identity is $0=(0,0)$ and the additive inverse of $a+bi$ is $-a-bi$. Multiplication distributes over addition, and $i^2=-1$, hence $(a+bi)(c+di)=ac+(ad+bc)i+bdi^2=(ac-bd)+(ad+bc)i$. The multiplicative identity is $1=(1,0)=1+0i$. If $z=a+bi$, then the *absolute value* of z is $|z|=\sqrt{a^2+b^2}$, which is equal to the length of the vector (a,b). Let $r=|a+bi|$. If θ is the angle determined by the vectors $z=a+bi$ and $1=(1,0)$, then the representation of z in polar coordinates is $z=a+bi=r\cos\theta+ir\sin\theta$. The complex conjugate of $z=a+bi$ is $\chi(z)=a-bi$. Then $z\chi(z)=a^2+b^2=|z|^2$ is a nonnegative real number. This implies if $z\neq 0$, then z is invertible and

$$z^{-1}=\frac{a-bi}{a^2+b^2}.$$

The power series for the functions e^x, $\cos x$, and $\sin x$ are

$$\begin{aligned}
e^x &= 1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\frac{x^5}{5!}+\frac{x^6}{6!}+\frac{x^7}{7!}+\frac{x^8}{8!}+\dots\\
\cos x &= 1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\frac{x^8}{8!}+\dots\\
\sin x &= x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\dots.
\end{aligned}$$

These power series converge for every real number x. We define e^{ix} to be the substitution of ix into the power series. Using the identities $i^2=-1$, $i^3=-i$, $i^4=1$, and $i^5=i$, we have

$$\begin{aligned}
e^{ix} &= 1+ix+\frac{i^2x^2}{2!}+\frac{i^3x^3}{3!}+\frac{i^4x^4}{4!}+\frac{i^5x^5}{5!}+\frac{i^6x^6}{6!}+\frac{i^7x^7}{7!}+\frac{i^8x^8}{8!}+\dots\\
&= 1+ix-\frac{x^2}{2!}-\frac{ix^3}{3!}+\frac{x^4}{4!}+\frac{ix^5}{5!}-\frac{x^6}{6!}-\frac{ix^7}{7!}+\frac{x^8}{8!}+\dots\\
&= \left(1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\frac{x^8}{8!}+\dots\right)+i\left(x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\dots\right)\\
&= \cos x+i\sin x.
\end{aligned}$$

Therefore, if $z=a+bi$ has polar representation $r\cos\theta+ir\sin\theta$, then the representation for z in exponential form is $a+bi=re^{i\theta}$.

PROPOSITION 1.4.3. *In exponential notation, arithmetic in $\mathbb{C}$ satisfies the following formulas.*

(1) (Additive inverse) $-re^{i\theta}=re^{i(\theta+\pi)}$.
(2) (Multiplication) $re^{i\theta}se^{i\phi}=(rs)e^{i(\theta+\phi)}$.
(3) (Complex conjugation) $\chi\left(re^{i\theta}\right)=re^{-i\theta}$.
(4) (Multiplicative inverse) $\left(re^{i\theta}\right)^{-1}=r^{-1}e^{-i\theta}$.
(5) (Square root) If $r\geq 0$, then $z^{1/2}=\sqrt{re^{i\theta}}=\sqrt{r}e^{i\theta/2}$.
(6) (nth root) If $r\geq 0$, then $z^{1/n}=\left(re^{i\theta}\right)^{1/n}=r^{1/n}e^{i\theta/n}$.

PROOF. The proof is left to the reader. □

CHAPTER 2

Groups

Groups arise in all areas of Mathematics. All of the other algebraic structures that arise are also based on groups. A module is an abelian group, a ring is an additive abelian group and the set of invertible elements of a ring is a group. For this reason the theorems of this chapter are fundamental.

In a concrete sense, a group is a set of permutations of a set. Galois first emphasized the importance of studying permutations of the roots of polynomials. Group Theory can be viewed as an axiomatic abstraction of permutation groups.

1. First properties of groups

The notion of a binary operation on a set was introduced in Section 1.1.5. The main ideas remain the same, but we recast them in light of the present context. Let G be a nonempty set with a binary operation $G \times G \to G$. Usually the binary operation on a group will be written multiplicatively or additively. In the multiplicative notation, an identity element will usually be denoted e or 1 and the inverse of an element a will be written a^{-1}. If additive notation is used, an identity is usually denoted 0 and $-a$ denotes the inverse of a.

1.1. Definitions and Terminology.

DEFINITION 2.1.1. Let G be a nonempty set with a multiplicative binary operation. If $a(bc) = (ab)c$ for all $a, b, c \in G$, then the binary operation is said to be associative. In this case, G is called a *semigroup*. If G is a semigroup and G contains an element e satisfying $ae = ea = a$ for all $a \in G$, then e is said to be an identity element and G is called a *monoid*. Let G be a monoid with identity element e. An element $a \in G$ is said to be *invertible* if there exists $a^{-1} \in G$ such that $aa^{-1} = a^{-1}a = e$. The element a^{-1} is called the *inverse of* a. A monoid in which every element is invertible is called a *group*. In other words, a group is a nonempty set G together with an associative binary operation such that an identity element e exists in G, and every element of G is invertible. If $xy = yx$ for all $x, y \in G$, then the binary operation is said to be commutative. A commutative group is called an *abelian group*.

If G has an additive binary operation, then the associative law is $(a+b)+c = a+(b+c)$ for all $a, b, c \in G$. The element $0 \in G$ is an identity element if $a + 0 = 0 + a = a$ for all $a \in G$. The element a is invertible if there exists an inverse element $-a \in G$ such that $a + (-a) = (-a) + a = 0$. The commutative law is $a + b = b + a$ for all $a, b \in G$. As a rule, additive notation is not used for nonabelian groups.

EXAMPLE 2.1.2. Let X be a nonempty set. A one-to-one correspondence $\sigma : X \to X$ is also called a permutation of X. The set of all permutations of X is denoted $\mathrm{Perm}(X)$. If σ and τ are permutations of X, then so is the composite function $\sigma\tau$, by Proposition 1.1.1. Therefore, $\mathrm{Perm}(X)$ is a group with identity element 1_X. If $|X| > 1$, then $\mathrm{Perm}(X)$ is nonabelian.

EXAMPLE 2.1.3. Here are some examples of abelian groups.

(1) Under addition, $\mathbb{Z}$ is an abelian group with identity 0. The inverse of x is written $-x$.
(2) Let $n \in \mathbb{N}$. Proposition 1.2.9 shows that under addition, $\mathbb{Z}/(n)$ is an abelian group with identity $[0]$. The inverse of $[x]$ is $[-x]$. We have $|\mathbb{Z}/(n)| = n$.
(3) Let $n \in \mathbb{N}$. Lemma 1.2.12 shows that the set of units modulo n, U_n, is a multiplicative abelian group. The identity element is $[1]$ and $|U_n| = \phi(n)$.

Let G be a multiplicative semigroup. The associative law on G says that $(ab)c = a(bc)$. In other words, a product of length three has a unique value regardless of how we associate the multiplications into binary operations using parentheses. When writing a product abc it is not necessary to use parentheses. The next lemma extends this result to products of arbitrary finite length.

LEMMA 2.1.4. *(General Associative Law) Let G be a semigroup, $n \geq 1$, and $x_1x_2\cdots x_n$ a product involving n elements of G. Then the product has a unique value regardless of how we associate the multiplications into binary operations using parentheses.*

PROOF. First we define a standard value for $x_1x_2\cdots x_n$ by the recursive formula:

$$x_1x_2\cdots x_n = \begin{cases} x_1 & \text{if } n = 1 \\ (x_1x_2\cdots x_{n-1})x_n & \text{if } n > 1. \end{cases}$$

Now we show that any association of $x_1x_2\cdots x_n$ will result in the value defined above. The proof is by induction on n. If $n \leq 3$, then this is true by the associative law on G. Inductively assume $n > 3$ and that the result holds for any product of length less than n. Let $x_1x_2\cdots x_n$ be a product involving n elements. Assume the product is associated into binary operations using parentheses. Then the last binary operation can be written as

$$(x_1x_2\cdots x_m)(x_{m+1}\cdots x_n)$$

and by the induction hypothesis, the two products $x_1x_2\cdots x_m$ and $x_{m+1}\cdots x_m$ have unique values regardless of how they are associated. If $m = n-1$, then we are done, by the induction hypothesis. Assume $1 \leq m < n-1$. Using the associative law on G and the induction hypothesis, we get

$$\begin{aligned}(x_1x_2\cdots x_m)(x_{m+1}\cdots x_n) &= (x_1x_2\cdots x_m)((x_{m+1}\cdots x_{n-1})x_n) \\ &= ((x_1x_2\cdots x_m)(x_{m+1}\cdots x_{n-1})x_n) \\ &= (x_1x_2\cdots x_{n-1})x_n) \\ &= x_1x_2\cdots x_n\end{aligned}$$

which completes the proof. □

DEFINITION 2.1.5. Let G be a group, $a \in G$, and n a nonnegative integer.

(1) If G is a multiplicative group, then the *nth power of a* is defined recursively by the formula:

$$a^n = \begin{cases} e & \text{if } n = 0 \\ aa^{n-1} & \text{if } n > 0. \end{cases}$$

We define a^{-n} to be $(a^{-1})^n$, which is equal to $(a^n)^{-1}$.
(2) If A and B are nonempty subsets of G, then

$$AB = \{xy \mid x \in A, y \in B\}.$$

(3) For an additive group G, the counterpart of the nth power is *left multiplication of a by n*, which is defined recursively by:

$$na = \begin{cases} 0 & \text{if } n = 0 \\ a + (n-1)a & \text{if } n > 0. \end{cases}$$

and $(-n)a$ is defined to be $n(-a)$, which is equal to $-(na)$.

(4) If A and B are nonempty subsets of the additive group G, then we define

$$A + B = \{x + y \mid x \in A, y \in B\}.$$

PROPOSITION 2.1.6. *Let G be a group and a, b, c elements of G.*

(1) There exists a unique x in G such that $ax = b$.
(2) There exists a unique y in G such that $ya = b$.
(3) We have $ab = ac$ if and only if $b = c$.
(4) We have $ab = cb$ if and only if $a = c$.

Parts (1) and (2) are called the solvability *properties, Parts (3) and (4) are called the* cancellation *properties.*

PROOF. (3): Assume we have $ab = ac$. Multiply both sides on the left by a^{-1} to get $a^{-1}ab = a^{-1}ac$. Since $a^{-1}ab = eb = b$ and $a^{-1}ac = ec = c$, we get $b = c$. Conversely, multiplying both sides of $b = c$ from the left with a yields $ab = ac$.

(1): Let $x = a^{-1}b$. Multiply by a on the left to get $ax = aa^{-1}b = eb = b$. If x' is another solution, then $ax = ax'$ and by Part (3) we have $x = x'$.

Parts (4) and (2) are proved in a similar manner. □

EXAMPLE 2.1.7. Let G be a group. Let $a \in G$ be a fixed element. Then "left multiplication by a" defines a function $\lambda_a : G \to G$, where $\lambda_a(x) = ax$. Part (1) of Proposition 2.1.6 says that λ_a is onto and Part (3) says that λ_a is one-to-one. Therefore, λ_a is a one-to-one correspondence. Likewise, "right multiplication by a" defines a one-to-one correspondence $\rho_a : G \to G$ where $\rho_a(x) = xa$.

DEFINITION 2.1.8. If G is a group, then the *order of G* is the cardinality of the underlying set. The order of G is denoted $[G : e]$ or $|G|$ or $o(G)$.

DEFINITION 2.1.9. Let G be a group and $a \in G$. The *order of a*, written $|a|$, is the least positive integer m such that $a^m = e$. If no such integer exists, then we say a has infinite order.

DEFINITION 2.1.10. Let G and G' be groups. A function $\theta : G \to G'$ is called an *isomorphism of groups*, if θ is a one-to-one correspondence and $\theta(xy) = \theta(x)\theta(y)$ for all $x, y \in G$. In this case, we say G and G' are *isomorphic* and write $G \cong G'$. From an abstract algebraic point of view, isomorphic groups are indistinguishable.

1.2. Examples of groups.

EXAMPLE 2.1.11. In this example we show that there is up to isomorphism only one group of order two. By Example 2.1.3, a group of order two exists, namely the additive group $\mathbb{Z}/2$. Let $G = \{e, a\}$ be an arbitrary group of order two, where e is the identity element. By Example 2.1.7, left multiplication by a is a permutation of G. Since $ae = a$, this implies $aa = e$. In other words, there is only one binary operation that makes $\{e, a\}$ into a group. If $G' = \{e, b\}$ is a group, then the function that maps $e \mapsto e$, $a \mapsto b$ is an isomorphism.

EXAMPLE 2.1.12. We know from Example 2.1.3 that the additive group $\mathbb{Z}/3$ is an abelian group of order three. In this example we show that up to isomorphism there is only one group of order three. Let $G = \{e, a, b\}$ be an arbitrary group of order three, where e is the identity element. By Example 2.1.7, λ_a and ρ_a are permutations of G. By cancellation, $ab = b$ leads to the contradiction $a = e$. Since $ae = a$, we conclude that $ab = e$ and $aa = b$. Similarly, $ba = b$ is impossible, hence we conclude that $ba = e$. We have shown that $G = \{e, a, a^2\}$ and a has order 3. Suppose $G' = \{e, c, c^2\}$ is another group of order 3. Then the assignments $a^i \mapsto c^i$ for $i = 0, 1, 2$ define an isomorphism.

EXAMPLE 2.1.13. If $X = \{x_1, \ldots, x_n\}$ is a finite set, then a binary operation on X can be represented as an n-by-n matrix with entries from X. Sometimes we call the matrix the "multiplication table" or "addition table". If the binary operation is $*$, then the entry in row i and column j of the associated matrix is the product $x_i * x_j$. For instance, the multiplication and addition tables for $\mathbb{Z}/6$ are:

$*$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

$+$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

If the binary operation $*$ on X is commutative, then the matrix is symmetric with respect to the main diagonal. If $X, *$ is a group, then by Example 2.1.7, each row of the multiplication table is a permutation of the top row and each column is a permutation of the leftmost column. See Exercise 2.1.28 for more examples.

EXAMPLE 2.1.14. Let $n \geq 1$ and $\mathbb{N}_n = \{1, 2, \ldots, n\}$. The set of all permutations of $\mathbb{N}_n$ is called the *symmetric group on n letters* and is denoted S_n. In Example 2.1.2 we saw that composition of functions makes $S_n = \mathrm{Perm}(X)$ into a group. As in Section 1.1.3, the group S_n has order $n!$. A permutation can be specified using an array of two rows. For example,

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & \ldots & n \\ a_1 & a_2 & a_3 & \ldots & a_n \end{bmatrix}$$

represents the permutation $\sigma(i) = a_i$. The so-called cycle notation is a very convenient way to represent elements of S_n. Let $\{a_1, \ldots, a_k\} \subseteq \mathbb{N}_n$. The *k-cycle* $\sigma = (a_1 a_2 \ldots a_k)$ is the permutation of $\mathbb{N}_n$ defined by:

$$\sigma(x) = \begin{cases} x & \text{if } x \notin \{a_1, \ldots, a_k\} \\ a_1 & \text{if } x = a_k \\ a_{i+1} & \text{if } x = a_i \text{ and } 1 \leq i < k. \end{cases}$$

Notice that a k-cycle has order k in the group S_n. The identity element of S_n is usually denoted e. For example, $(abc)(ab) = (ac)$ and $(ab)(abc) = (bc)$. Therefore, S_n is nonabelian if $n > 2$. The group table for $S_3 = \{e, (abc), (acb), (ab), (ac), (bc)\}$ is:

$*$	e	(abc)	(acb)	(ab)	(ac)	(bc)
e	e	(abc)	(acb)	(ab)	(ac)	(bc)
(abc)	(abc)	(acb)	(e)	(ac)	(bc)	(ab)
(acb)	(acb)	(e)	(abc)	(bc)	(ab)	(ac)
(ab)	(ab)	(bc)	(ac)	(e)	(acb)	(abc)
(ac)	(ac)	(ab)	(bc)	(abc)	(e)	(acb)
(bc)	(bc)	(ac)	(ab)	(acb)	(abc)	(e)

EXAMPLE 2.1.15. Let T be a regular triangle with vertices labeled $1,2,3$. A *symmetry* of T is any transformation $\sigma : T \to T$ that preserves distances. Therefore, σ is a permutation of the three vertices. Conversely, a permutation of $\{1,2,3\}$ uniquely determines a symmetry of T. The group of symmetries of T is therefore equal to S_3.

EXAMPLE 2.1.16. Now let $n > 2$ and let T_n be a regular n-gon with vertices labeled $1,2,\ldots,n$ consecutively. A symmetry of T_n is any transformation $\sigma : T_n \to T_n$ that preserves distances. Therefore, σ is a permutation of the n vertices. If $n > 3$, a permutation of $\{1,2,\ldots,n\}$ does not necessarily determine a symmetry of T_n. When $n > 3$, the group of symmetries of T_n is therefore a proper subgroup of S_n. The group of all symmetries of T_n is called the *dihedral group* D_n. A rotation of T_n through an angle of $2\pi/n$ corresponds to the permutation

$$R = \begin{bmatrix} 1 & 2 & 3 & \ldots & n-1 & n \\ 2 & 3 & 4 & \ldots & n & 1 \end{bmatrix}$$

which in cycle notation is the n-cycle $R = (12\ldots n)$. Therefore, R^k is a rotation of T_n through an angle of $2\pi k/n$, hence R has order n. A top to bottom flip of T_n across the line of symmetry containing vertex 1 corresponds to the permutation defined by

$$H = \begin{cases} \begin{bmatrix} 1 & 2 & 3 & \ldots & k & k+1 & \ldots & n-1 & n \\ 1 & n & n-1 & \ldots & k+2 & k+1 & \ldots & 3 & 2 \end{bmatrix} & \text{if } n = 2k \text{ is even,} \\ \begin{bmatrix} 1 & 2 & 3 & \ldots & k & k+1 & \ldots & n-1 & n \\ 1 & n & n-1 & \ldots & k+1 & k & \ldots & 3 & 2 \end{bmatrix} & \text{if } n = 2k-1 \text{ is odd.} \end{cases}$$

In cycle notation, H can be represented as

$$H = \begin{cases} (2,n)(3,n-1)\cdots(k,k+2) & \text{if } n = 2k \text{ is even,} \\ (2,n)(3,n-1)\cdots(k,k+1) & \text{if } n = 2k-1 \text{ is odd.} \end{cases}$$

Then $HH = e$, hence H has order 2. The reader should verify that $HRH = R^{-1}$. Any symmetry of T_n is either a rotation or a flip followed by a rotation. Therefore we see that $D_n = \{H^i R^j \mid 0 \le i \le 1, 0 \le j < n\}$ is a nonabelian group of order $2n$.

EXAMPLE 2.1.17. Let R_4 be a nonsquare rectangle with vertices labeled consecutively $1,2,3,4$. The group of symmetries of R_3 can be viewed as a subgroup of S_4 as well as a subgroup of D_4. In the notation of Example 2.1.16, the group of symmetries of R_4 is $\{H^i R^j \mid 0 \le i \le 1, 0 \le j \le 1\}$, which is a group of order four. In cycle notation, this group is $\{e,(14)(23),(12)(34),(13)(24)\}$. Note that the group is abelian and every element satisfies the identity $x^2 = e$.

EXAMPLE 2.1.18. The *quaternion* 8*-group* is $Q_8 = \{1,-1,i,-i,j,-j,k,-k\}$ with identity element 1. The multiplication rules are: $(-1)^2 = 1$, $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$. This is an example of a nonabelian group of order eight. For a continuation of this example, see Exercise 2.4.19.

EXAMPLE 2.1.19. Let F be a field. If α is a nonzero element of F, then α has a multiplicative inverse, denoted α^{-1}. The set of all nonzero elements of F is a multiplicative group. This group is denoted F^* and is called the *group of units of* F.

EXAMPLE 2.1.20. Let F be a field. The set of all n-by-n matrices over F is denoted $M_n(F)$. In this example, we assume the reader is familiar with the basic properties for multiplication of matrices. In particular, multiplication of matrices is associative. That is, if $\alpha,\beta,\gamma \in M_n(F)$, then $(\alpha\beta)\gamma = \alpha(\beta\gamma)$. We will not include the tedious but elementary proof of this fact here. Instead, we mention that in Corollary 4.5.7 below a general proof will be given that matrix multiplication is associative and distributes over matrix addition. In this example our goal is to show that the set of 2-by-2 matrices over F with nonzero determinant is a group. For $n = 2$, the determinant function $\det : M_2(F) \to F$ is defined by

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

To show that the determinant function is multiplicative, start with the product of two arbitrary 2-by-2 matrices:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae+bg & af+bh \\ ce+dg & cf+dh \end{pmatrix}.$$

The determinant formula applied on the left hand side yields: $(ad-bc)(eh-fg) = adeh - adfg - bceh + bcfg$. The reader should verify that this is equal to the determinant of the right hand side: $(ae+bg)(cf+dh) - (ce+dg)(af+bh)$. A matrix α is invertible if there is a matrix β such that $\alpha\beta = \beta\alpha = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Taking determinants, this implies $\det\alpha \det\beta = 1$. In other words, if α is invertible, then $\det\alpha \neq 0$. Notice that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} = (ad-bc) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

If $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc \neq 0$, then the matrix is invertible and the inverse is given by the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = (ad-bc)^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

The set

$$\mathrm{GL}_2(F) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(F) \mid ad - bc \neq 0 \right\}$$

is the set of all invertible 2-by-2 matrices over F and is called the *general linear group of* 2*-by-*2 *matrices over* F. For a continuation of this example when $F = \mathbb{Z}/2$ is the field of order 2, see Exercise 2.1.26.

EXAMPLE 2.1.21. The Klein Viergruppe, or 4-group, is $V = \{e,a,b,c\}$ with multiplication rules: $a^2 = b^2 = c^2 = e$, $ab = ba = c$. Notice that V is isomorphic to the group of symmetries of a nonsquare rectangle presented in Example 2.1.17 by the mapping: $a \mapsto (14)(23)$, $b \mapsto (12)(34)$, $c \mapsto (13)(24)$.

1.3. Exercises.

EXERCISE 2.1.22. Let G be a monoid with identity element e.

(1) Show that G has exactly one identity element. In other words, show that if $e' \in G$ has the property that $ae' = e'a = a$, then $e = e'$.

(2) Show that an invertible element of G has a unique inverse. In other words, if $aa^{-1} = a^{-1}a = e$ and $aa' = a'a = e$, then $a^{-1} = a'$.
(3) Suppose $a, r, \ell \in G$ satisfy the identities: $ar = e$ and $\ell a = e$. Show that $r = \ell$ and a is invertible.
(4) Suppose every element of G has a left inverse. In other words, assume for every $a \in G$ there exists $a_l \in G$ such that $a_l a = e$. Show that G is a group.
(5) If $a \in G$ is invertible, then a^{-1} is invertible and $(a^{-1})^{-1} = a$.
(6) If a and b are invertible elements of G, then ab is invertible and $(ab)^{-1} = b^{-1}a^{-1}$.

EXERCISE 2.1.23. Let G be a group and $x, y \in G$. Prove the following:

(1) If $x^2 = x$, then $x = e$. We say that a group has exactly one idempotent.
(2) If $xy = e$, then $y = x^{-1}$.
(3) $(x^{-1})^{-1} = x$.
(4) $(xy)^{-1} = y^{-1}x^{-1}$.

EXERCISE 2.1.24. Let G be a group. The *opposite group of* G is denoted G^o. As a set, G^o is equal to G. The binary operation on G^o is reversed from that of R. Writing the multiplication of R by juxtaposition and multiplication of R^o with the asterisk symbol, we have $x * y = yx$. Show that G^o is a group. Show that G is isomorphic to G^o. (Hint: Show that the function defined by $x \mapsto x^{-1}$ is an isomorphism from G to G^o.)

EXERCISE 2.1.25. Let G be a group. Prove the following:

(1) If $x^2 = e$ for all $x \in G$, then G is abelian.
(2) If $|G| = 2n$ for some $n \in \mathbb{N}$, then there exists $x \in G$ such that $a \neq e$ and $a^2 = e$.

EXERCISE 2.1.26. In this example, we assume the reader is familiar with the basic properties for multiplication of matrices. In particular, multiplication of matrices is associative and the product of a two-by-two matrix times a two-by-one column vector is defined by:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} au + bv \\ cu + dv \end{pmatrix}.$$

Let $G = \mathrm{GL}_2(\mathbb{Z}/2)$ be the group of two-by-two invertible matrices over the field $\mathbb{Z}/2$ (see Example 2.1.20). List the elements of G and construct the group table (see Example 2.1.13). Show that G has two elements of order three and three elements of order two. Let

$$a = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, c = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and consider the set of column vectors $\{a, b, c\}$ over $\mathbb{F}_2$. For every matrix α in G, show that left multiplication by the matrix α defines a permutation of the set $\{a, b, c\}$. Comparing the group table for G with the group table given in Example 2.1.14 for S_3, the symmetric group on 3 letters, show that $\mathrm{GL}_2(\mathbb{Z}/2)$ is isomorphic to S_3.

EXERCISE 2.1.27. Let K and H be groups. Define a binary operation on $K \times H$ by $(x_1, y_1)(x_2, y_2) = (x_1x_2, y_1y_2)$. Show that this makes $K \times H$ into a group with identity element (e, e), and the inverse of (x, y) is (x^{-1}, y^{-1}). Show that $K \times H$ is abelian if and only if K and H are both abelian.

EXERCISE 2.1.28. For various values of n, each of the following matrices is an n-by-n multiplication table representing a binary operation $*$ on the set $I_n = \{0, 1, \ldots, n-1\}$. In each case, determine whether the binary operation (a) is commutative, (b) is associative, (c) has an identity element, and (d) is a group.

(1)

$*$	0	1	2	3
0	0	0	0	0
1	0	1	1	3
2	0	2	3	0
3	0	3	1	2

(2)

$*$	0	1	2	3	4	5	6	7
0	4	2	6	0	7	1	5	3
1	5	4	0	1	6	7	3	2
2	1	7	4	2	5	3	0	6
3	0	1	2	3	4	5	6	7
4	7	6	5	4	3	2	1	0
5	6	0	3	5	2	4	7	1
6	2	3	7	6	1	0	4	5
7	3	5	1	7	0	6	2	4

(3)

$*$	0	1	2	3	4	5	6	7
0	4	5	3	2	0	1	7	6
1	7	4	5	6	1	2	3	0
2	3	7	4	0	2	6	5	1
3	2	6	0	4	3	7	1	5
4	0	1	2	3	4	5	6	7
5	6	0	1	7	5	3	2	4
6	5	3	7	1	6	0	4	2
7	1	2	6	5	7	4	0	3

(4)

$*$	0	1	2	3	4	5	6	7
0	7	2	1	4	3	6	5	0
1	2	7	0	5	6	3	4	1
2	1	0	7	6	5	4	3	2
3	4	5	6	7	0	1	2	3
4	3	6	5	0	7	2	1	4
5	6	3	4	1	2	7	0	5
6	5	4	3	2	1	0	7	6
7	0	1	2	3	4	5	6	7

(5)

$*$	0	1	2
0	2	0	1
1	0	1	2
2	1	2	0

(6)

$*$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	0	4	5	3
2	2	0	1	5	3	4
3	3	5	4	0	2	1
4	4	3	5	1	0	2
5	5	4	3	2	1	0

2. Subgroups and cosets

2.1. First properties of subgroups.

DEFINITION 2.2.1. If G is a group and H is a nonempty subset of G that is a group under the binary operation on G, then we say H is a *subgroup of* G and write $H \leq G$.

LEMMA 2.2.2. *Let G be a group and H a nonempty subset of G. The following are equivalent.*

(1) H is a subgroup of G.
(2) For all a, b in H we have $ab \in H$ and $a^{-1} \in H$.
(3) For all a, b in H we have $ab^{-1} \in H$.

PROOF. (2) implies (1): Let $a \in H$. Then $e = aa^{-1} \in H$. The associative law applies on G, hence on H. The other group properties are included in (2).

(1) implies (3): Let a and b be elements of H. If H is a group, then $b^{-1} \in H$ and $ab^{-1} \in H$.

(3) implies (2): Let a and b be elements of H. By (3) we have $aa^{-1} = e \in H$, $ea^{-1} = a^{-1} \in H$, and $a(b^{-1})^{-1} = ab \in H$. □

EXAMPLE 2.2.3. Let G be a group. Then $\{e\}$ and G are both subgroups of G. We call these the *trivial subgroups of* G. A nontrivial subgroup is also called a *proper subgroup*.

PROPOSITION 2.2.4. *Let G be a group and H a finite subset of G. If for all $a, b \in H$ we have $ab \in H$, then H is a subgroup of G.*

PROOF. Assume $a,b \in H$ implies $ab \in H$. By Lemma 2.2.2, to show H is a subgroup it suffices to show that $a \in H$ implies $a^{-1} \in H$. Let $|H| = n$. Define $f : \mathbb{N}_{n+1} \to H$ be defined by $f(i) = a^i$. Since $a \in H$, we see from Definition 2.1.5 that f is well defined. The Pigeonhole Principle (Exercise 1.1.11) implies that there exists a pair $0 < i < j \leq n+1$ such that $a^i = a^j$. Then $j - i > 0$, so $e = a^{j-i}$ is in H. If $j - i = 1$, then $a = e$, which implies $a^{-1} = e \in H$. If $j - i > 1$, then $e = a^{j-i} = aa^{j-i-1}$, which implies $a^{-1} = a^{j-i-1} \in H$. □

LEMMA 2.2.5. *Let G be a group and $X \subseteq G$. Let $\mathscr{S} = \{H \leq G \mid X \subseteq H\}$, and let*

$$\langle X \rangle = \bigcap_{H \in \mathscr{S}} H$$

be the intersection of all subgroups of G containing X. Then the following are true.

(1) $\langle X \rangle$ is the smallest subgroup of G containing X.

(2) $\langle X \rangle$ is the trivial subgroup $\{e\}$ if $X = \emptyset$, otherwise

$$\langle X \rangle = \left\{ x_1^{e_1} \cdots x_n^{e_n} \mid n \geq 1, e_i \in \mathbb{Z}, x_i \in X \right\}.$$

PROOF. (1): We know $\mathscr{S}$ is nonempty because $G \in \mathscr{S}$. Therefore, (1) follows straight from Exercise 2.2.21.

(2): If $X = \emptyset$, then $\{e\} \in \mathscr{S}$, so $\langle X \rangle = \{e\}$. Assume $X \neq \emptyset$. By Lemma 2.2.2 (1), the set $S = \left\{ x_1^{e_1} \cdots x_n^{e_n} \mid n \geq 1, e_i \in \mathbb{Z}, x_i \in X \right\}$ is a subgroup of G. Since $X \subseteq S$, we have $\langle X \rangle \subseteq S$. Let $x_1^{e_1} \cdots x_n^{e_n}$ be a typical element of S. For each i, $x_i \in X$ implies x_i is in the group $\langle X \rangle$. By Definition 2.1.5, the power $x_i^{e_i}$ is in $\langle X \rangle$. Therefore, the product $x_1^{e_1} \cdots x_n^{e_n}$ is in $\langle X \rangle$. This proves $S \subseteq X$. □

DEFINITION 2.2.6. In the context of Lemma 2.2.5, the set $\langle X \rangle$ is called the *subgroup of G generated by X*. If $X = \{x_1, \ldots, x_n\}$ is a finite subset of G, then we sometimes write $\langle X \rangle$ in the form $\langle x_1, \ldots, x_n \rangle$. A subgroup $H \leq G$ is said to be *finitely generated* if there exists a finite subset $\{x_1, \ldots, x_n\} \subseteq H$ such that $H = \langle x_1, \ldots, x_n \rangle$. We say H is *cyclic* if $H = \langle x \rangle$ for some $x \in H$.

DEFINITION 2.2.7. Let G be a group and H a subgroup of G. If x and y are elements of G, then we say *x is congruent to y modulo H* if $x^{-1}y \in H$. In this case we write $x \equiv y \pmod{H}$.

LEMMA 2.2.8. *Let G be a group and H a subgroup. Then congruence modulo H is an equivalence relation on G.*

PROOF. If $x \in G$, then $x^{-1}x = e \in H$, so $x \equiv x \pmod{H}$. Assume $x \equiv y \pmod{H}$. Then $x^{-1}y \in H$, which implies $y^{-1}x = (x^{-1}y)^{-1} \in H$, hence $y \equiv x \pmod{H}$. Assume $x \equiv y \pmod{H}$ and $y \equiv z \pmod{H}$. Then $x^{-1}yy^{-1}z = x^{-1}z \in H$, which implies $x \equiv z \pmod{H}$. □

LEMMA 2.2.9. *Let G be a group, H a subgroup, and $x, y \in G$. The following are equivalent.*

(1) $x \equiv y \pmod{H}$.

(2) $y = xh$ for some $h \in H$.

(3) $xH = yH$.

PROOF. (1) is equivalent to (2): We have $x \equiv y \pmod{H}$ if and only if $x^{-1}y \in H$ which is true if and only if $x^{-1}y = h$ for some $h \in H$ which is equivalent to $y = xh$ for some $h \in H$.

(3) implies (2): We have $y = ye \in yH = xH$. Therefore, $y = xh$ for some $h \in H$.

(2) implies (3): Suppose $y = xh$, for some $h \in H$. For every $z \in H$, $yz = x(hz) \in xH$. Hence $yH \subseteq xH$. Also, $x = yh^{-1}$ implies $xz = y(h^{-1}z) \in yH$, which implies $xH \subseteq yH$. □

2.2. Cosets and Lagrange's Theorem. Let G be a group and H a subgroup. By Lemma 2.2.8, congruence modulo H is an equivalence relation on G. Therefore G is partitioned into equivalence classes. If $x \in G$, then by Lemma 2.2.9, the equivalence class of x is $xH = \{y \in G \mid y = xh \text{ for some } h \in H\}$. The set xH is called *the left coset of x modulo H*. The set of all left cosets of G modulo H is $G/H = \{xH \mid x \in G\}$. By Proposition 1.1.2 two cosets are either disjoint or equal as sets. The *index of H in G* is the cardinality of the set G/H and is denoted $[G : H]$.

There is a right hand version of the above, which we will briefly describe here. We say *x is right congruent to y modulo H* if $yx^{-1} \in H$. This defines an equivalence relation on G. The equivalence class of x is the set Hx which is called the *right coset of x modulo H*. The set of all right cosets is denoted $H\backslash G$. In general, the partitions G/H and $H\backslash G$ are not equal. That is, a left coset is not necessarily a right coset (see Lemma 2.3.4). In Exercise 2.2.23 the reader is asked to show that there is a one-to-one correspondence between G/H and $H\backslash G$.

LEMMA 2.2.10. *Let G be a group and $H \leq G$. Given $x, y \in G$ there is a one-to-one correspondence $\phi : xH \to yH$ defined by $\phi(z) = (yx^{-1})z$. If $|H|$ is finite, then all left cosets of H have the same number of elements.*

PROOF. For any $h \in H$, $yx^{-1}xh = yh \in yH$. We see that ϕ is a well defined function. The function $\psi(w) = xy^{-1}w$ is the inverse to ϕ. □

If H is a subgroup of G, then *a complete set of left coset representatives for H in G* is a subset $\{a_i \mid i \in I\}$ of G where we have exactly one element from each left coset. The index set I can be taken to be G/H. If $\{a_i \mid i \in I\}$ is a complete set of left coset representatives, then $G = \cup_{i \in I} a_i H$ is a partition of G. For example, if $m \geq 1$, then Proposition 1.2.9 (2) shows that $\{0, 1, \ldots, m-1\}$ is a complete set of left coset representatives for $\langle m \rangle$ in $\mathbb{Z}$.

THEOREM 2.2.11. *If $K \leq H \leq G$, then $[G : K] = [G : H][H : K]$. If two of the three indices are finite, then so is the third.*

PROOF. Let $\{a_i \mid i \in I\}$ be a complete set of left coset representatives for H in G and Let $\{b_j \mid j \in J\}$ be a complete set of left coset representatives for K in H. Then $G = \cup_{i \in I} a_i H$ is a partition of G and $H = \cup_{j \in J} b_j K$ is a partition of H. So

$$\begin{aligned} G &= \bigcup_{i \in I} a_i H \\ &= \bigcup_{i \in I} a_i \left(\bigcup_{j \in J} b_j K \right) \\ &= \bigcup_{i \in I} \left(\bigcup_{j \in J} a_i b_j K \right). \end{aligned}$$

To finish the proof, we show that $\{a_i b_j \mid (i, j) \in I \times J\}$ is a complete set of left coset representatives for K in G. It suffices to show the cosets $a_i b_j K$ are pairwise disjoint. Assume $a_i b_j K = a_s b_t K$. Then $a_i b_j = a_s b_t k$ for some $k \in K$. Recall that b_j, b_t, k are in H. Then we have $a_i = a_s h$, for some $h \in H$. Hence $a_i H = a_s H$, which implies $i = s$. Canceling, we get $b_j = b_t k$, or $b_j K = b_t K$, which implies $j = t$. This proves $[G : K] = [G : H][H : K]$. The index $[G : K]$ is infinite if and only if $[G : H]$ is infinite or $[H : K]$ is infinite. This proves the theorem. □

COROLLARY 2.2.12. *(Lagrange's Theorem) If G is a group and $H \leq G$, then $|G| = [G:H]|H|$.*

PROOF. Apply Theorem 2.2.11 with $K = \langle e \rangle$. □

2.3. A counting theorem.

LEMMA 2.2.13. *Let G be a group containing subgroups H and K. Then HK is a subgroup of G if and only if $HK = KH$.*

PROOF. See Definition 2.1.5 (2) for the definition of the set HK. First assume $HK = KH$. To show HK is a subgroup we show that the criteria of Lemma 2.2.2 (1) are satisfied. In the following, h, h_1, h_2, h_3 denote elements of H and k, k_1, k_2, k_3 denote elements of K. Let h_1k_1 and h_2k_2 be arbitrary elements of HK. Since $HK = KH$, there exist h_3, k_3 such that $k_1h_2 = h_3k_3$. Now $(h_1k_1)(h_2k_2) = h_1(k_1h_2)k_2 = h_1(h_3k_3)k_2 = (h_1h_3)(k_3k_2)$ is an element of HK. By Exercise 2.1.23, $(hk)^{-1} = k^{-1}h^{-1}$ is is an element of $KH = HK$. This proves HK is a subgroup.

Conversely, suppose HK is a subgroup. Consider the function $i : G \to G$ defined by $i(x) = x^{-1}$. By Exercise 2.1.23, i^2 is the identity function. Thus i is a one-to-one correspondence. Since HK is a group, the restriction of i to HK is a one-to-one correspondence. That is, $i(HK) = HK$. If $hk \in HK$, then $i(hk) = (hk)^{-1} = k^{-1}h^{-1}$ is in KH, which shows $HK = i(HK) \subseteq KH$. Consider $kh \in KH$. Then $i(kh) = (kh)^{-1} = h^{-1}k^{-1}$ is in HK. Therefore, kh is the inverse of an element in the subgroup HK. By Lemma 2.2.2, $kh \in HK$, which implies $KH \subseteq HK$. □

THEOREM 2.2.14. *Let G be a group. If H and K are finite subgroups of G, then*

$$|HK| = \frac{|H||K|}{|H \cap K|}.$$

PROOF. We do not assume HK is a group. Let $C = H \cap K$. Then C is a subgroup of H. Let $\{h_1, \ldots, h_n\}$ be a full set of left coset representatives of C in H, where $n = [H:C]$. Then $H = \cup_{i=1}^{n} h_iC$ is a disjoint union. Since $C \subseteq K$ we have $CK = K$, hence

$$HK = \bigcup_{i=1}^{n} h_iCK = \bigcup_{i=1}^{n} h_iK.$$

The last union is a disjoint union. To see this, suppose $h_iK = h_jK$. Then $h_j^{-1}h_i \in H \cap K = C$, which implies $i = j$. By Lemma 2.2.10 we can now count the cardinality of HK:

$$|HK| = \sum_{i=1}^{n} |K| = n|K| = [H : H \cap K]|K|.$$

By Corollary 2.2.12, we are done. □

2.4. Cyclic subgroups. In the next theorem we show that the additive group $\mathbb{Z}$ is cyclic and every subgroup is of the form $\langle n \rangle$ for some $n \geq 0$. Moreover, the equivalence relation of Definition 2.2.7 defined in terms of the subgroup $\langle n \rangle$ is equal to the equivalence relation of Definition 1.2.8 defined in terms of divisibility by n.

THEOREM 2.2.15. *Let $\mathbb{Z}$ be the additive group of integers.*

(1) Every subgroup of $\mathbb{Z}$ is cyclic. The trivial subgroups of $\mathbb{Z}$ are: $\langle 0 \rangle$ and $\mathbb{Z} = \langle 1 \rangle$. If H is a nontrivial subgroup, then there is a unique $n > 1$ such that $H = \langle n \rangle = n\mathbb{Z} = \{nk \mid k \in \mathbb{Z}\}$.

(2) If $n \geq 1$ and $H = \langle n \rangle$, then $x \equiv y \pmod{H}$ if and only if $x \equiv y \pmod{n}$. That is, the coset $x + \langle n \rangle$ in $\mathbb{Z}/\langle n \rangle$ is equal to the congruence class $[x]$ in $\mathbb{Z}/n$.

PROOF. Let $H \leq \mathbb{Z}$ and assume $H \neq \langle 0 \rangle$. If $x \in H - \langle 0 \rangle$, then so is $-x$. By the Well Ordering Principle (Axiom 1.2.1) there is a least positive integer in H, say n. We prove that $H = n\mathbb{Z}$. Let $x \in H$. By the Division Algorithm (Proposition 1.2.3) we can write $x = nq + r$ where $0 \leq r < n$. By Definition 2.1.5, $nq \in H$. Therefore, $r = x - nq$ is in H. By the choice of n, this implies $r = 0$. Hence $x \in n\mathbb{Z}$. □

Let G be a group and a an element of finite order in G. Recall (Definition 2.1.9) that the order of a, written $|a|$, is the least positive integer m such that $a^m = e$.

LEMMA 2.2.16. *Let G be a group, $a \in G$, and assume $|a| = m$ is finite. Then the following are true.*

(1) $|a| = |\langle a \rangle|$.
(2) $\langle a \rangle = \{e, a, a^2, \ldots, a^{m-1}\}$.
(3) For each $n \in \mathbb{Z}$, $a^n = e$ if and only if m divides n.
(4) For each $n \in \mathbb{Z}$, $|a^n| = m/\gcd(m,n)$.
(5) Let $b \in G$. Assume $|b| = n$ is finite, $ab = ba$, and $\langle a \rangle \cap \langle b \rangle = \langle e \rangle$. Then $|ab| = \operatorname{lcm}(m,n)$.

PROOF. (1) and (2): Let $m = |a|$. Then $m > 0$, $a^m = e$, and if $m > 1$, then $a^{m-1} \neq e$. Let $n \in \mathbb{Z}$. Applying Proposition 1.2.3, there exist unique integers q and r such that $n = mq + r$ and $0 \leq r < m$. Then $a^n = (a^m)qa^r = a^r$. Therefore, $\langle a \rangle = \{e, a, a^2, \ldots, a^{m-1}\}$. It follows that $|\langle a \rangle| = m$.

(3): First assume $n = mq$. Then we have $a^{mq} = (a^m)^q = e^q = e$. Conversely assume $a^n = e$. By Parts (1) and (2), if $n = mq + r$ and $0 \leq r < m$, then $a^r = e$, which implies $r = 0$.

(4) and (5): This part of the proof is Exercise 2.2.27. □

COROLLARY 2.2.17. *If $|G|$ is finite, and $a \in G$, then the following are true.*

(1) $|a|$ is finite.
(2) $|a|$ divides $|G|$.
(3) $a^{|G|} = e$.

PROOF. (1): Proposition 2.2.4 shows that $|a|$ is finite.

(2) and (3): These follow immediately from Lemma 2.2.16 and Corollary 2.2.12. □

COROLLARY 2.2.18. *Let $a \in \mathbb{Z}$. Then the following are true.*

(1) (Euler) If $m \in \mathbb{N}$ and $\gcd(a,m) = 1$, then $a^{\phi(m)} \equiv 1 \pmod{m}$.
(2) (Fermat) If p is prime and $\gcd(a,p) = 1$, then $a^{p-1} \equiv 1 \pmod{p}$.

PROOF. As noted in Example 2.1.3, U_n, the group of units modulo n, has order $\phi(n)$. If p is prime, then $\phi(p) = p - 1$. □

COROLLARY 2.2.19. *Let G be a group satisfying $|G| > 1$. If G has no proper subgroup, then $|G|$ is finite, $|G|$ is prime, and G is cyclic.*

PROOF. Let $a \in G - \langle e \rangle$. Since G has no proper subgroup and $\langle e \rangle \neq \langle a \rangle$ is a subgroup of G, we have $\langle a \rangle = G$. Look at the set $S = \{e, a, a^2, \ldots\}$. If there is a relation of the form $a^k = a^m$, where $k < m$, then $|a|$ is finite, hence G is finite. Conversely, if G is finite, then Proposition 2.2.4 shows that there is a relation $a^k = a^m$, where $k < m$. Assume for contradiction's sake that G is infinite. Then $a \neq a^n$, for all $n > 1$. Thus, $\langle a^2 \rangle$ is a proper subgroup of G, a contradiction. We conclude that $G = \langle a \rangle = \{e, a, \ldots, a^{n-1}\}$ is a finite

cyclic group of order n, for some n. Assume for contradiction's sake that $n = xy$ where $1 < x \leq y < n$. By Lemma 2.2.16 (4), $\langle a^x \rangle = \{e, a^x, a^{2x}, \ldots, a^{(y-1)x}\}$ has order y, hence G has a proper subgroup, which is a contradiction. This proves n is prime. □

COROLLARY 2.2.20. *Let G be a group. If G has only a finite number of subgroups, then G is finite.*

PROOF. Suppose G is an infinite group. We prove that G has infinitely many subgroups. Let $x_1 \in G$ and set $X_1 = \langle x_1 \rangle$. By Theorem 2.2.15, the additive group of integers $\mathbb{Z}$ has infinitely many distinct subgroups, namely $\{\langle n \rangle \mid n \geq 0\}$. If X_1 is infinite, then the same proof shows that X_1 has infinitely many distinct subgroups, namely $\{\langle x_1^n \rangle \mid n \geq 0\}$. From now on assume every element of G has finite order. Then $G - \langle x_1 \rangle$ is infinite. Pick $x_2 \in G - \langle x_1 \rangle$. Then $\langle x_1 \rangle \neq \langle x_2 \rangle$. Assume inductively that $n \geq 1$ and $x_1, x_2, \ldots, x_n$ are in G such that $X_1 = \langle x_1 \rangle, \ldots, X_n = \langle x_n \rangle$ are n distinct subgroups. Then $\cup_{i=1}^n X_i$ is finite. Pick $x_{n+1} \in G - X_1 - X_2 - \cdots - X_n$ and set $X_{n+1} = \langle x_{n+1} \rangle$. Then by induction there exists an infinite collection $\{X_i \mid i \geq 1\}$ of distinct subgroups of G. □

2.5. Exercises.

EXERCISE 2.2.21. (An intersection of subgroups is a subgroup.) Let G be a group, I a nonempty set, and $\{H_i \mid i \in I\}$ a family of subgroups of G indexed by I. Show that

$$\bigcap_{i \in I} H_i$$

is a subgroup of G.

EXERCISE 2.2.22. Let G be a group and X, Y, Z subgroups of G. Prove that if $Y \subseteq X$, then $X \cap YZ = Y(X \cap Z)$.

EXERCISE 2.2.23. Let G be a group and H a subgroup of G. We denote by G/H the set of all left cosets of H in G, and by $H\backslash G$ the set of all right cosets of H in G. Show that the assignment $xH \mapsto Hx^{-1}$ defines a one-to-one correspondence between G/H and $H\backslash G$.

EXERCISE 2.2.24. Let G be a group containing finite subgroups H and K. If $|H|$ and $|K|$ are relatively prime, show that $H \cap K = \langle e \rangle$.

EXERCISE 2.2.25. This exercise is a continuation of Exercise 2.1.27. Let K and H be groups and $K \times H$ the product group. Show that $\{(x, e) \mid x \in K\}$ and $\{(e, y) \mid y \in H\}$ are subgroups of $K \times H$.

EXERCISE 2.2.26. Consider the symmetric group S_3 of order 6. Show that S_3 has 4 proper subgroups. Let H be the subgroup of order 2 generated by the transposition (12). Compute the three left cosets of H and the three right cosets of H.

EXERCISE 2.2.27. Prove Parts (4) and (5) of Lemma 2.2.16.

EXERCISE 2.2.28. Let p be a prime number and G a finite group of order p. Prove:

(1) G has no proper subgroup.
(2) There exists $a \in G$ such that $G = \langle a \rangle$.

EXERCISE 2.2.29. Let $(\mathbb{R}, +)$ denote the additive group on $\mathbb{R}$. Then $(\mathbb{Q}, +)$ is a subgroup of $(\mathbb{R}, +)$ and $(\mathbb{Z}, +)$ is a cyclic subgroup of both $(\mathbb{Q}, +)$ and $(\mathbb{R}, +)$. Show that the set $\{x \in \mathbb{R} \mid 0 \leq x < 1\}$ is a complete set of left coset representatives for $\mathbb{Z}$ in $\mathbb{R}$. Show that the set $\{x \in \mathbb{Q} \mid 0 \leq x < 1\}$ is a complete set of left coset representatives for $\mathbb{Z}$ in $\mathbb{Q}$. See Exercise 2.3.21 for a continuation of this exercise.

3. Homomorphisms and normal subgroups

3.1. Definition and first properties of normal subgroups. A function from one group to another that preserves the binary operations is called a homomorphism. If H is a subgroup of G, then H is a normal subgroup if and only if the binary operation on G turns the set of left cosets G/H into a group and in this case the natural map $G \to G/H$ is a homomorphism of groups (Lemma 2.3.4).

DEFINITION 2.3.1. A *homomorphism of groups* is a function $\phi : G \to G'$ from a group G to a group G' such that $\phi(xy) = \phi(x)\phi(y)$ for all $x, y \in G$. If ϕ is onto, we say ϕ is an *epimorphism*. If ϕ is one-to-one, we say ϕ is a *monomorphism*. If ϕ is one-to-one and onto, then as in Definition 2.1.10 we say ϕ is an *isomorphism*. A homomorphism from G to G is called an *endomorphism of G*. An isomorphism from G to G is called an *automorphism of G*.

DEFINITION 2.3.2. Let $\phi : G \to G'$ be a homomorphism of groups. The *kernel of ϕ* is $\ker(\phi) = \{x \in G \mid \phi(x) = e\}$.

DEFINITION 2.3.3. Let G be a group. For every $a \in G$, let $\alpha_a : G \to G$ be defined by $\alpha_a(x) = a^{-1}xa$. If X is a nonempty subset of G, then $\alpha_a(X) = a^{-1}Xa$ is called the *conjugate of X by a*.

The next lemma lists the fundamental properties of normal subgroups. The definition follows the lemma.

LEMMA 2.3.4. *Let G be a group and H a subgroup of G. The following are equivalent.*

(1) For each $x \in G$, $x^{-1}Hx \subseteq H$.
(2) For each $x \in G$, $x^{-1}Hx = H$.
(3) For each $x \in G$ there exists $y \in G$ such that $xH = Hy$.
(4) For each $x \in G$, $xH = Hx$.
(5) For each $x \in G$ and $y \in G$, $xHyH = xyH$.
(6) There is a well defined binary operation $G/H \times G/H \to G/H$ on G/H defined by the rule $(xH, yH) \mapsto xyH$.
(7) There is a binary operation on G/H such that the natural map $\eta : G \to G/H$ is a homomorphism of groups.
(8) There exists a group G' and a homomorphism of groups $\theta : G \to G'$ such that $H = \ker\theta$.

PROOF. (1) implies (2): Let $x \in G$. First apply (1) to x, yielding $x^{-1}Hx \subseteq H$. Now conjugate by x^{-1} and apply (1) with x^{-1} to get $H = (xx^{-1})H(xx^{-1}) \subseteq xHx^{-1} \subseteq H$.

(2) implies (3): Let $x \in G$. Apply (2) to x^{-1} to get $xHx^{-1} = H$. This implies $xH = Hx$.

(3) implies (4): Given $x \in G$, there exists $y \in G$ such that $xH = Hy$. Since x is in $xH = Hy$, this implies $x = hy$ for some $h \in H$. Therefore $y = h^{-1}x$ and $Hy = Hh^{-1}x = Hx$.

(4) implies (5): Let $x \in G$ and $y \in G$. By (4) applied to y, $yH = Hy$. Therefore, $xHyH = x(Hy)H = x(yH)H = xyH$.

(5) implies (6): This is immediate.

(6) implies (7): By (6), $(xH, yH) \mapsto xyH$ defines a binary operation on G/H. The associative law on G implies the associative law also holds on G/H. The identity element is the coset eH and $(xH)^{-1} = x^{-1}H$. Therefore G/H is a group and it is now clear that the natural map $\eta : G \to G/H$ is a homomorphism.

(7) implies (8): The kernel of $\eta : G \to G/H$ is $\eta^{-1}(eH) = H$.

(8) implies (1): Let $\theta : G \to G'$ be a homomorphism of groups and assume $H = \ker\theta$. By Exercise 2.3.15, the preimage of a subgroup of G' is a subgroup of G. Therefore,

$\ker(\theta) = \theta^{-1}(\langle e \rangle)$ is a subgroup of G. Given $x \in G$ and $h \in H$ we have $\theta(h) = e$. Hence $\theta(x^{-1}hx) = \theta(x)^{-1}\theta(h)\theta(x) = \theta(x)^{-1}\theta(x) = e$. Therefore, $x^{-1}Hx \subseteq \ker\theta = H$. □

DEFINITION 2.3.5. If G is a group and H is a subgroup of G satisfying any of the equivalent conditions of Lemma 2.3.4, then we say H is a *normal subgroup of* G. The group G/H is called the *quotient group*, or *factor group*. If N is a normal subgroup of G, we sometimes write $N \trianglelefteq G$.

EXAMPLE 2.3.6. Let G be a group.

(1) The trivial subgroups $\langle e \rangle$ and G are normal in G.
(2) If G is abelian and H is a subgroup of G, then for every $x \in G$, $xH = Hx$, hence H is normal. The quotient group G/H is abelian because G is abelian.

3.2. The Isomorphism Theorems. The Fundamental Theorem of Group Homomorphisms, Theorem 2.3.11, says that any homomorphism of groups $\theta : A \to B$ factors in a natural way into a surjection $A \to A/\ker(\theta)$ followed by an injection $A/\ker(\theta) \to B$. This proves us with a valuable tool for defining a homomorphism on a quotient group A/N. As applications, we prove the Isomorphism Theorems (Theorem 2.3.12) and the Correspondence Theorem (Theorem 2.3.13).

LEMMA 2.3.7. *Let $\phi : G \to G'$ and $\phi_1 : G' \to G''$ be homomorphisms of groups. Then the following are true.*

(1) The composite $\phi_1\phi : G \to G''$ is a homomorphism of groups.
(2) The kernel of ϕ, $\ker(\phi)$, is a normal subgroup of G.
(3) The function ϕ is one-to-one if and only if $\ker(\phi) = \langle e \rangle$.

PROOF. (1): This follows straight from: $\phi_1\phi(xy) = \phi_1(\phi(x)\phi(y)) = \phi_1\phi(x)\phi_1\phi(y)$.
(2): By Lemma 2.3.4 (8), $\ker(\phi)$ is a normal subgroup of G.
(3): If ϕ is one-to-one, then $\ker(\phi) = \phi^{-1}(\langle e \rangle) = \langle e \rangle$. If $\ker(\phi) = \langle e \rangle$ and $\phi(x) = \phi(y)$, then $\phi(xy^{-1}) = \phi(x)\phi(y)^{-1} = e$, so $xy^{-1} \in \ker(\phi)$. Therefore, $x = y$ and ϕ is one-to-one. □

EXAMPLE 2.3.8. If $\phi : G \to G'$ is an isomorphism of groups, then as in Definition 2.1.10 we say G is isomorphic to G', and write $G \cong G'$. If $\phi_1 : G' \to G''$ is another isomorphism of groups, then by Lemma 2.3.7 and Exercise 1.1.9, the composite $\phi_1\phi$ is an isomorphism. The reader should verify that isomorphism defines an equivalence relation on the set of all groups.

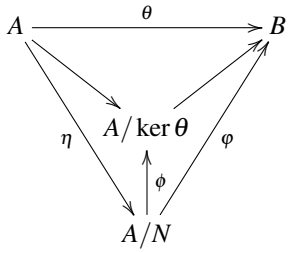
EXAMPLE 2.3.9. Let G be a group. The set of all automorphisms of G is denoted $\mathrm{Aut}(G)$. By Lemma 2.3.7 the composition of automorphisms is an automorphism. In the notation of Example 2.1.2, $\mathrm{Aut}(G)$ is a subgroup of $\mathrm{Perm}(G)$.

EXAMPLE 2.3.10. Let G be a group and $a \in G$. Then conjugation by a defines the function $\alpha_a : G \to G$, where $\alpha_a(x) = a^{-1}xa$. In Exercise 2.3.19 the reader is asked to prove that α_a is an automorphism of G. We call α_a the inner automorphism of G defined by a. The set of all inner automorphisms is a subgroup of $\mathrm{Aut}(G)$.

THEOREM 2.3.11. *(Fundamental Theorem of Group Homomorphisms) Let $\theta : A \to B$ be a homomorphism of groups. Let N be a normal subgroup of A contained in $\ker\theta$. There exists a homomorphism $\varphi : A/N \to B$ satisfying the following.*

(a) $\varphi(aN) = \theta(a)$, or in other words $\theta = \varphi\eta$.
(b) φ is the unique homomorphism from $A/N \to B$ such that $\theta = \varphi\eta$.
(c) $\mathrm{im}\,\theta = \mathrm{im}\,\varphi$.

(d) $\ker\varphi = \eta(\ker\theta) = \ker(\theta)/N$.
(e) φ *is one-to-one if and only if* $N = \ker\theta$.
(f) φ *is onto if and only if* θ *is onto.*
(g) There is a unique epimorphism $\phi : A/N \to A/\ker\theta$ *such that the diagram*



commutes.

PROOF. The map φ exists by Exercise 1.1.13. The proofs of (a) – (f) are left as an exercise for the reader. Part (g) results from an application of Parts (a) – (f) to the natural map $A \to A/\ker\theta$. □

THEOREM 2.3.12. *(The Isomorphism Theorems) Let G be a group.*

(a) If $\theta : G \to G'$ *is a homomorphism of groups, then the map* $\varphi : G/\ker\theta \to \operatorname{im}\theta$ *sending the coset* $x\ker\theta$ *to* $\theta(x)$ *is an isomorphism of groups.*
(b) If A and B are subgroups of G and B is normal, then natural map

$$\frac{A}{A\cap B} \to \frac{AB}{B}$$

sending the coset $x(A\cap B)$ *to the coset xB is an isomorphism of groups.*
(c) If A and B are normal subgroups of G and $A \subseteq B$*, then* B/A *is a normal subgroup of* G/A *and the natural map*

$$\frac{G/A}{B/A} \to G/B$$

sending the coset containing xA to the coset xB is an isomorphism of groups.

PROOF. (a): By Exercise 2.3.15, the image of G is a subgroup of G'. This is Parts (e) and (f) of Theorem 2.3.11.

(b): By Exercise 2.3.18, AB is a group, B is normal in AB, and $A\cap B$ is normal in A. Let $f : A \to (AB)/B$ be the set containment map $A \to AB$ followed by the natural map $AB \to (AB)/B$. If $a \in A$ and $b \in B$, then $abB = aB$, hence f is onto. Let $a \in A$. Then $aB = B$ if and only if $a \in B$. Therefore the kernel of f is $A\cap B$. Part (b) follows from Part (a) applied to the homomorphism f.

(c): By Theorem 2.3.11 (g) applied to the natural map $G \to G/B$, there is a natural epimorphism $\phi : G/A \to G/B$ defined by $\phi(xA) = xB$. The kernel of ϕ consists of those cosets xA such that $x \in B$. That is, $\ker\phi = B/A$. Part (c) follows from Part (a) applied to the homomorphism ϕ. □

THEOREM 2.3.13. *(The Correspondence Theorem) Let G be a group and A a normal subgroup of G. There is a one-to-one order-preserving correspondence between the subgroups B such that* $A \subseteq B \subseteq G$ *and the subgroups of* G/A *given by* $B \mapsto B/A$*. Moreover, B is a normal subgroup of G if and only if* B/A *is a normal subgroup of* G/A*.*

PROOF. Let $\eta : G \to G/A$ be the natural homomorphism. By Exercise 2.3.15, if B is a subgroup of G, then $\eta(B)$ is a subgroup of G/A, and if H is a subgroup of G/A, then $\eta^{-1}(H)$ is a subgroup of G containing A. If $B_1 \subseteq B_2$, then $\eta(B_1) \subseteq \eta(B_2)$. Likewise, if $H_1 \subseteq H_2$, then $\eta^{-1}(H_1) \subseteq \eta^{-1}(H_2)$. Since η is onto, $\eta\eta^{-1}(H) = H$. By Exercise 2.3.15, if B is a subgroup of G containing A, then $B = \eta^{-1}\eta(B)$. This proves the first claim.

For the last claim, let B be a subgroup of G containing A. If B is normal, then by Theorem 2.3.12 (c), $\eta(B)$ is normal in G/A. Conversely assume $\eta(B)$ is normal in G/A. Then B is equal to the kernel of the composite map $G \to G/A \to (G/A)/\eta(B)$, hence is normal in G. □

EXAMPLE 2.3.14. Let $(\mathbb{R},+)$ be the additive abelian group of real numbers and $(\mathbb{R}_{>0},\cdot)$ the multiplicative abelian group of positive real numbers. Define $\phi : (\mathbb{R},+) \to (\mathbb{R}_{>0},\cdot)$ by $\phi(x) = e^x$. Then $\phi(x+y) = e^{x+y} = e^x e^y = \phi(x)\phi(y)$, so ϕ is a homomorphism. Define $\psi : (\mathbb{R}_{>0},\cdot) \to (\mathbb{R},+)$ by $\psi(x) = \ln x$. Then $\psi(xy) = \ln xy = \ln x + \ln y = \psi(x) + \psi(y)$, so ψ is a homomorphism. Since ϕ and ψ are inverses of each other, they are isomorphisms. Hence $(\mathbb{R},+)$ and $(\mathbb{R}_{>0},\cdot)$ are isomorphic groups.

3.3. Exercises.

EXERCISE 2.3.15. Let $f : G \to G'$ be a homomorphism of groups. Prove:

(1) $f(e) = e$.
(2) $f(x^{-1}) = f(x)^{-1}$.
(3) If H is a subgroup of G, then $f(H)$ is a subgroup of G'. If there is a containment relation $H_1 \subseteq H_2$, then $f(H_1) \subseteq f(H_2)$.
(4) If H' is a subgroup of G', then $f^{-1}(H')$ is a subgroup of G and $\ker f$ is a subgroup of $f^{-1}(H')$. If there is a containment relation $H_1' \subseteq H_2'$, then $f^{-1}(H_1') \subseteq f^{-1}(H_2')$.
(5) If H is a subgroup of G and $\ker f \subseteq H$, then $f^{-1}f(H) = H$.
(6) If G is abelian, then $\mathrm{im}(f)$ is abelian.

EXERCISE 2.3.16. Let $G,+$ be an additive abelian group. Let $n \in \mathbb{Z}$ and $x \in G$. If $n > 0$, then $nx = \sum_{i=1}^{n} x = x + \cdots + x$ is the sum of n copies of x. If $n < 0$, then $nx = |n|(-x) = \sum_{i=1}^{|n|}(-x)$, and $0x = 0$.

(1) Show that "left multiplication by n" defines a function $\lambda_n : G \to G$ by the rule $\lambda_n(x) = nx$. Show that λ_n is an endomorphism of G.
(2) Show that the kernel of λ_n is $G(n) = \{x \in G \mid |x| \mid n\}$, hence $G(n)$ is a subgroup of G.
(3) Show that the image of λ_n is $nG = \{nx \mid x \in G\}$, hence nG is a subgroup of G.

When the group operation is written multiplicatively, the counterpart of λ_n is the "nth power map" which is denoted $\pi^n : G \to G$ and is defined by $\pi^n(x) = x^n$. In this case, $\mathrm{im}(\pi^n)$ is denoted G^n.

EXERCISE 2.3.17. Let G be a group and H a subgroup. Prove that if $[G : H] = 2$, then H is a normal subgroup.

EXERCISE 2.3.18. Let G be a group containing subgroups H, K, and N. Prove the following:

(1) If N is a normal subgroup of G, then NK is a subgroup of G. Moreover, K is a subgroup of NK, and N is a normal subgroup of NK.
(2) If N is normal, then $N \cap H$ is a normal subgroup of H.
(3) If H and K are both normal, then HK is a normal subgroup of G.

EXERCISE 2.3.19. Let G be a group. For every $a \in G$, let $\alpha_a : G \to G$ be defined by $\alpha_a(x) = a^{-1}xa$. In the terminology of Definition 2.3.3, $\alpha_a(x)$ is the conjugate of x by a. Prove that α_a is an automorphism of G.

EXERCISE 2.3.20. (The conjugate of a subgroup is a subgroup.) Let G be a group, S a nonempty subset of G, and $a \in G$. The *conjugate of S by a* is defined to be $S^a = a^{-1}Sa$. Prove that S is a subgroup of G if and only if S^a is a subgroup of G.

EXERCISE 2.3.21. Let $S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$. Then $S^1 = \{e^{2\pi i\theta} \mid 0 \leq \theta < 1\}$ is the unit circle in the complex plane (see Section 1.4).

(1) Show that multiplication in $\mathbb{C}$ makes S^1 into a group.
(2) Let $(\mathbb{R}, +)$ denote the additive group on $\mathbb{R}$. Show that the function $f : (\mathbb{R}, +) \to S^1$ defined by $f(\theta) = e^{2\pi i\theta}$ is an onto homomorphism. Compute the kernel of f. Show that f induces an isomorphism $\mathbb{R}/\mathbb{Z} \cong S^1$ (see Exercise 2.2.29).
(3) If $n \in \mathbb{N}$, then the nth power map $z \mapsto z^n$ is an endomorphism of S^1 (see Exercise 2.3.16). Let μ_n denote the kernel of the nth power map. Show that $\mu_n = \{e^{2\pi ik/n} \mid k \in \mathbb{Z}\}$ is the set of all nth roots of unity in $\mathbb{C}$.
(4) Show that the function $\phi : \mathbb{Z} \to \mu_n$ defined by $\phi(k) = e^{2\pi ik/n}$ is an epimorphism. Compute the kernel of ϕ. Show that ϕ induces an isomorphism $\mathbb{Z}/n \cong \mu_n$.
(5) Let $\mu = \cup_{n\geq 1}\mu_n$. Show that μ is a group. Define $h : \mathbb{Q} \to \mu$ by $h(r) = e^{2\pi ir}$. Show that h is an epimorphism. Compute the kernel of h. Show that h induces an isomorphism $\mathbb{Q}/\mathbb{Z} \cong \mu$ (see Exercise 2.2.29).

EXERCISE 2.3.22. Let G be a finite group of order $n = [G : e]$. Let p be a prime number such that $p \mid n$ and $p^2 > n$. Assume G contains a subgroup H of order p. (This is always true, by Cauchy's Theorem, Theorem 2.7.3.) Prove:

(1) H is the unique subgroup of G of order p.
(2) H is a normal subgroup of G.

EXERCISE 2.3.23. A group G is said to be *simple* if the only normal subgroups of G are $\langle e\rangle$ and G. Prove that a group G is simple if and only if for every nontrivial homomorphism of groups $f : G \to G'$, f is a monomorphism.

EXERCISE 2.3.24. This exercise is a continuation of Exercise 2.2.25. Let K and H be groups and $K \times H$ the product group. Define four functions

(1) $\iota_1 : K \to K \times H$, $\iota_1(x) = (x, e)$
(2) $\iota_2 : H \to K \times H$, $\iota_2(y) = (e, y)$
(3) $\pi_1 : K \times H \to K$, $\pi_1(x, y) = x$
(4) $\pi_2 : K \times H \to H$, $\pi_2(x, y) = y$

Show that ι_1 and ι_2 are monomorphisms. Show that π_1 and π_2 are epimorphisms. Show that $\operatorname{im}\iota_1 = \ker\pi_1 = K \times \{e\}$ and $\operatorname{im}\iota_2 = \ker\pi_2 = \{e\} \times H$.

3.4. More on Cyclic groups. A cyclic group $A = \langle a\rangle$ is generated by a single element. Theorem 2.3.25 shows that if A is infinite, then A is isomorphic to the additive group $\mathbb{Z}$. In this case A has two generators, namely a, and a^{-1}. If A is finite of order n, then A is isomorphic to $\mathbb{Z}/n$ and A has $\phi(n)$ generators, namely $\{a^i \mid 1 \leq i \leq n-1, \gcd(i, n) = 1\}$. Lemma 2.3.26 shows that any homomorphism $A \to G$ of groups defined on A is completely determined by the image of a generator. Necessary and sufficient conditions for the existence of a homomorphism $A \to G$ are derived. In Theorem 2.3.27 we show that the group of all automorphisms of a cyclic group of order n is isomorphic to the group of units modulo n. The group of automorphisms of an infinite cyclic group is a group of order two. As

an application of these theorems on cyclic groups, we exhibit the classic proof by mathematical induction that a finite abelian group of order n contains an element of order p if p is a prime divisor of n (Theorem 2.3.28).

THEOREM 2.3.25. *(Fundamental Theorem on Cyclic Groups) Let $A = \langle a \rangle$ be a cyclic group. Then the following are true.*

(1) A is abelian.
(2) Every subgroup of A is cyclic.
(3) Every homomorphic image of A is cyclic.
(4) There is a unique $n \geq 0$ such that A is isomorphic to $\mathbb{Z}/\langle n \rangle$.
(5) If $n = 0$, then
  (a) A is infinite and
  (b) A is isomorphic to $\mathbb{Z}$.
(6) If $n > 0$, then
  (a) A isomorphic to $\mathbb{Z}/n$, hence A is finite of order n,
  (b) if H is a subgroup of A, then $|H|$ divides n,
  (c) for every positive divisor d of n, A has a unique subgroup of order d, namely $\langle a^{n/d} \rangle$,
  (d) if d is a positive divisor of n, then A has $\phi(d)$ elements of order d, where ϕ is the Euler function.

PROOF. (4): Let $\theta : \mathbb{Z} \to A$ be the function defined by $\theta(i) = a^i$. Since A is generated by a, θ is onto, by Lemma 2.2.5. Since $\theta(i+j) = a^{i+j} = a^i a^j = \theta(i)\theta(j)$, θ is an epimorphism. By Theorem 2.2.15 there is a unique $n \geq 0$ such that $\ker(\theta) = \langle n \rangle$. By Theorem 2.3.12 (1), θ induces an isomorphism $\bar{\theta} : \mathbb{Z}/\langle n \rangle \to A$.

(1): This follows from (4) and Exercise 2.3.15 (6).

(2) and (3) and (5): These follow from (4) and Theorems 2.2.15 and 2.3.13.

(6): Assume $n > 0$ and d is a positive divisor of n. By Lemma 2.2.16, $|a^{n/d}| = d$. Thus, $\langle a^{n/d} \rangle$, is a subgroup of order d. Now suppose $|a^x| = d$. By Lemma 2.2.16, $\gcd(x,n) = n/d$. By Bézout's Identity, Lemma 1.2.5, we can write $n/d = xu + nv$, for some $u, v \in \mathbb{Z}$. Since $a^{n/d} = (a^x)^u (a^n)^v = (a^x)^u$ we see that $\langle a^{n/d} \rangle \subseteq |a^x| = d$. Both groups have order d, hence they are equal. By Lemma 2.2.16, the number of elements of order n in A is equal to the cardinality of the set $\{x \in \mathbb{Z} \mid 1 \leq x \leq n \text{ and } \gcd(x,n) = 1\}$, which is equal to $\phi(n)$. Therefore, the number of elements of order d in a cyclic group of order d is $\phi(d)$. □

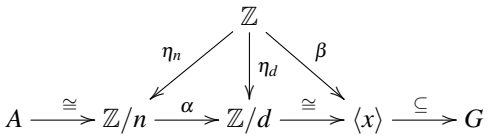
LEMMA 2.3.26. *Let $A = \langle a \rangle$ be a cyclic group and G any group.*

(1) Let $\phi : A \to G$ be a homomorphism of groups. Then ϕ is completely determined by the value $\phi(a)$.
(2) Let $x \in G$.
  (a) If the order of A is infinite, then there is a homomorphism $\theta : A \to G$ defined by $\theta(a) = x$.
  (b) If A has finite order $|A| = n$, then there is a homomorphism $\theta : A \to G$ defined by $\theta(a) = x$ if and only if x has finite order $|x| = d$ and $d \mid n$.

PROOF. (1): We have $\phi(a^i) = \phi(a)^i$.

(2): Part (a) was proved in the proof of Part (4) of Theorem 2.3.25. We prove Part (b). Assume A is finite and $|A| = n$. If there is a homomorphism $\theta : A \to G$, then by Exercise 2.3.40 the order of $\theta(a)$ is a divisor of n. Conversely, assume $|x| = d < \infty$ and $d \mid n$. By Theorem 2.3.25 there is an isomorphism $A \cong \mathbb{Z}/n$ defined by $a^i \mapsto [i]$ and a commutative

diagram



where $\beta(1) = x$, η_n and η_d are the natural maps, and α exists by Exercise 1.2.19. The homomorphism θ is the composition of the four homomorphisms in the bottom row. □

THEOREM 2.3.27. *Let $n \in \mathbb{N}$ be a positive integer. The group of automorphisms of the cyclic group of order n is isomorphic to the group of units modulo n. That is,*

$$\mathrm{Aut}(\mathbb{Z}/n) \cong U_n$$

which is a group of order $\phi(n)$. The group of automorphisms of the infinite cyclic group $\mathbb{Z}$ is isomorphic to the group of order two. That is,

$$\mathrm{Aut}(\mathbb{Z}) \cong \{1, -1\}.$$

PROOF. We utilize Theorem 2.3.25, Lemma 2.3.26, and Exercise 2.3.16. Let $A = \langle a \rangle$. Given $r \in \mathbb{Z}$, the rth power map on A is denoted $\pi^r : A \to A$ and is defined by $\pi^r(a) = a^r$. If $\alpha : A \to A$ is an endomorphism of A, then $\alpha(a) = a^s$ for some integer s. Since

$$\alpha(a^t) = \alpha(a)^t = (a^s)^t = a^{st} \tag{3.1}$$

we see that $\alpha = \pi^s$. That is, every endomorphism of A is π^r for some $r \in \mathbb{Z}$. This also shows $\pi^s \pi^t = \pi^{st}$. The image of π^r is the subgroup $\langle a^r \rangle$.

Case 1: Assume A is finite of order n. Then $a^r = a^s$ if and only if $r \equiv s \pmod{n}$. This proves there are n distinct endomorphisms of A, namely $\{\pi^0, \pi^1, \ldots, \pi^{n-1}\}$. The generators of A are $\{a^r \mid \gcd(r,n) = 1\}$, which is a set of order $\phi(n)$. Since π^r is one-to-one and onto if and only if a^r is a generator of A, this proves that there are $\phi(n)$ automorphisms of A, namely $\{\pi^r \mid 1 \leq r \leq n-1, \gcd(r,n) = 1\}$. By Example 2.1.3, the group of units modulo n is an abelian group of of order $\phi(n)$. Define $\theta : \mathrm{Aut}(\mathbb{Z}/n) \to U_n$ by $\theta(\pi^r) = r$. Then we have shown that θ is an isomorphism of groups.

Case 2: Assume A is infinite. Then $a^r = a^s$ if and only if $r = s$. By Theorem 2.2.15, the two generators of A are $\{a, a^{-1}\}$. Therefore, the two automorphisms of A are π^1 and π^{-1}. □

In general, if G is a finite group and p is a prime divisor of $|G|$, then G has an element of order p. This is known as Cauchy's Theorem and we will eventually present two proofs in Corollary 2.4.14 and Theorem 2.7.3. As an application of Theorem 2.3.25, an abelian version of Cauchy's Theorem is stated and proved in Theorem 2.3.28 below. The proof is by induction on the order of G. The induction step uses Lagrange's Theorem (Corollary 2.2.12) and the fact that if N is a subgroup of G, then G/N is an abelian group (Example 2.3.6). The key step in the induction argument is that an element of order p in the quotient group G/N "lifts" to an element in G whose order is a multiple of p.

THEOREM 2.3.28. *(Cauchy's Theorem for Abelian Groups) Let G be a finite abelian group and p a prime number. If p divides $|G|$, then G contains an element of order p.*

PROOF. The proof is by induction on the order of G. Let $n = |G|$. Since p divides n, we know $n > 1$. If $p = |G|$, then by Exercise 2.2.28, there exists $a \in G$ such that $G = \langle a \rangle$, hence $|a| = p$. Inductively assume n is composite and that the result holds for all abelian groups of order less than n. By Corollary 2.2.19, we know G has a proper subgroup, call it N. If p divides $|N|$, then by our induction hypothesis, N has an element of order p.

Therefore, assume p does not divide $|N|$. Since G is abelian, by Example 2.3.6, N is a normal subgroup and G/N is abelian. By Corollary 2.2.12, p divides $|N|[G:N]$. Since p does not divide $|N|$, we have p divides $[G:N]$. By our induction hypothesis, G/N has an element of order p. Suppose $b \in G$ and bN has order p in G/N. Since G is finite, b has finite order. By Exercise 2.3.40, p divides the order of b. By Theorem 2.3.25, $\langle b \rangle$ contains an element of order p. □

EXAMPLE 2.3.29. In this example we show that up to isomorphism there are exactly two groups of order six. By Example 2.1.3, we know that $\mathbb{Z}/6$ is an abelian group of order six. We know from Example 2.1.14 that the symmetric group on 3 letters, S_3, is a nonabelian group of order 6. Let G be a group of order six. Let $a \in G$ and set $A = \langle a \rangle$. By Corollary 2.2.17, $|a| \in \{1,2,3,6\}$. If G has an element of order 6, then by Theorem 2.3.25, G is isomorphic to $\mathbb{Z}/6$. Assume from now on that G has no element of order 6. For contradiction's sake, suppose G has no element of order 3. Then every element of G satisfies $x^2 = e$. By Exercise 2.1.25, G is abelian and there exists $a \in G$ such that $|a| = 2$. Then $A = \langle a \rangle$ is normal and G/A has order three. By Exercise 2.3.40, if the generator of G/A is bA, then b has order 3 or 6, a contradiction. We have shown that G has an element a of order 3. If $A = \langle a \rangle$, then by Exercise 2.3.22, A is the unique subgroup of order 3. Then $G - A$ consists of elements of order 2. Let $b \in G - A$. The coset decomposition of G is $A \cup bA = \{e, a, a^2\} \cup \{b, ba, ba^2\}$. Since $[G:A] = 2$, by Exercise 2.3.17 A is normal. By Lemma 2.3.4, $bA = Ab$. Therefore, $ab \in \{b, ab, a^2b\}$. We know $ab \neq b$ since $a \neq e$. If $ba = ab$, then by Lemma 2.2.16, $|ab| = 6$, a contradiction. Therefore, $ab = a^2b$. We have proved that $G = \{e, a, a^2, b, ab, a^2b\}$ where $a^3 = b^2 = e$ and $ab = a^2b$. The reader should verify that the assignments $a \mapsto (123)$, $a^2 \mapsto (132)$, $b \mapsto (12)$, $ab \mapsto (13)$, and $a^2b \mapsto (23)$ define an isomorphism $G \cong S_3$.

3.5. The center of a group. The center of a group is defined and as an exercise the reader is asked to prove that the center is a normal subgroup. As examples, we compute the center of the quaternion 8-group, the dihedral groups, the symmetric groups, and the general linear group of 2-by-2 matrices over a field.

DEFINITION 2.3.30. Let G be a group. The *center of* G, denoted $Z(G)$, is defined to be $\{x \in G \mid xa = ax \text{ for all } a \in G\}$. In Exercise 2.3.38 the reader is asked to prove that $Z(G)$ is a normal subgroup of G.

EXAMPLE 2.3.31. Let Q_8 be the quaternion 8-group of Example 2.1.18. In Exercise 2.4.19 the reader is asked to prove that the center of Q_8 is the unique subgroup of order two.

EXAMPLE 2.3.32. Let $n \geq 3$ and let D_n be the dihedral group (see Example 2.1.16). Then D_n is the group of symmetries of a regular n-gon. If H is the horizontal flip and R the rotation, then $D_n = \{H^i R^j \mid 0 \leq i \leq 1, 0 \leq j < n\}$ is a nonabelian group of order $2n$. The relations $H^2 = R^n = e$ and $HRH = R^{-1}$ hold. Hence the conjugate of R by H is R^{-1}. We show that if $n = 2k$ is even, then $Z(D_n)$ is the subgroup of order two generated by R^k. Conjugation by H is an automorphism, so if $0 < i < n$, then $HR^iH = R^{-i}$. We see that R^i is in $Z(D_n)$ if and only if $R^i = R^{-i}$, which is true if and only if $i = 0$ or $n = 2k$ is even and $i = k$. It follows that the center of $D_n = \langle e \rangle$ if n is odd. In summary, we have shown that

$$Z(D_n) = \begin{cases} \langle R^{n/2} \rangle & \text{if } n \text{ is even} \\ \langle e \rangle & \text{if } n \text{ is odd.} \end{cases}$$

EXAMPLE 2.3.33. Let $n \geq 3$ and let S_n be the symmetric group on n letters (see Example 2.1.14). We show that $Z(S_n) = \langle e \rangle$. Let $\pi \in S_n$ and assume $\pi \neq e$. First assume $\pi(a) = b$ and $\pi(b) = c$, where a, b, c are distinct. Let τ be the 2-cycle (ab). Then $\pi\tau(a) = \pi(b) = c$ and $\tau\pi(a) = \tau(b) = a$, which shows π is not central. Now suppose $\pi(a) = b$ and $\pi(b) = a$. Let σ be the 2-cycle (bc), where a, b, c are distinct. Then $\pi\sigma(a) = \pi(a) = b$ and $\sigma\pi(a) = \sigma(b) = c$, which shows π is not central. If $\pi \neq e$, then π falls into one of these two cases. This shows $Z(S_n) = \langle e \rangle$.

EXAMPLE 2.3.34. Let F be a field and $\mathrm{GL}_n(F)$ the general linear group of invertible n-by-n matrices over F. For instance, if $n = 1$, then $\mathrm{GL}_1(F)$ is simply the set $F - \{0\}$ of invertible elements in F, which we denote F^*. If $n = 2$, then

$$\mathrm{GL}_2(F) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid ad - bc \neq 0 \right\}.$$

To compute the center, assume $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a central matrix. Then

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} d & c \\ b & a \end{pmatrix}$$

shows that $a = d$ and $b = c$. Now

$$\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a-b & 0 \\ b & a+b \end{pmatrix}$$

shows that $b = 0$. Therefore, a central matrix is diagonal. It is routine to show that a diagonal matrix $\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$ is central. This computation shows that $Z(\mathrm{GL}_2(F))$ is equal to $\left\{ \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \mid a \in F^* \right\}$. If we define $\delta : F^* \to \mathrm{GL}_2(F)$ to be the diagonal map, $\delta(x) = \begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}$, then δ is a monomorphism and $\mathrm{im}(\delta) = Z(\mathrm{GL}_2(F))$. The quotient, $\mathrm{GL}_2(F)/F^*$, is denoted $\mathrm{PGL}_2(F)$ and is called the *projective general linear group of* 2*-by-*2 *matrices over* F.

EXAMPLE 2.3.35. Let F be a field. Let $\det : \mathrm{GL}_2(F) \to F^*$ be the determinant function, where $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$. In Example 2.1.20 we showed that det is an epimorphism on multiplicative groups. This is proved in Lemma 4.6.5 below for all n. The kernel, $\ker(\det)$, which is the set of all matrices with determinant equal to 1, is denoted $SL_2(F)$ and is called the *special linear group of* 2*-by-*2 *matrices over* F. By Theorem 2.3.12 (a) there is an isomorphism of groups

$$\mathrm{GL}_2(F)/\mathrm{SL}_2(F) \cong F^*.$$

See Exercise 2.5.15 for a computation of $SL_2(\mathbb{Z}/3)$.

EXAMPLE 2.3.36. As in Example 2.1.14, the group of permutations of the set $\{1, 2, 3\}$ is

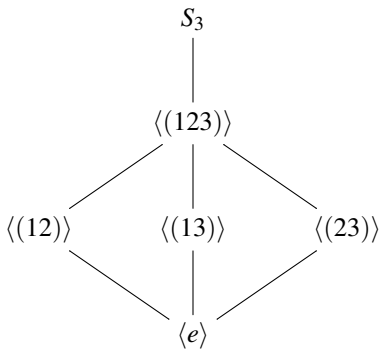
$$S_3 = \{e, (123), (132), (12), (13), (23)\}$$

and is called the *symmetric group on* 3 *elements*. The group S_3 is isomorphic to D_3, the group of symmetries of an equilateral triangle (see Example 2.1.15). Also, S_3 is isomorphic to $\mathrm{GL}_2(\mathbb{Z}/2)$, the group of invertible 2-by-2 matrices over the field of order 2 (see

Exercise 2.1.26). The group table for S_3 is listed in Example 2.1.14. The cyclic subgroups of S_3 are:

$$\begin{gathered}\langle e\rangle = \{e\}\\ \langle (123)\rangle = \langle (132)\rangle = \{e,(123),(132)\}\\ \langle (12)\rangle = \{e,(12)\}\\ \langle (13)\rangle = \{e,(13)\}\\ \langle (23)\rangle = \{e,(23)\}\end{gathered}$$

Since S_3 is a subgroup of itself, there are exactly 6 subgroups. The center of S_3 is the trivial subgroup $\langle e\rangle$, by Example 2.3.33. The commutator subgroup (see Exercise 2.3.42) of S_3 is the cyclic subgroup $\langle (123)\rangle$, by Exercise 2.3.43. There is one subgroup of order 6, one subgroup of order 3, three subgroups of order 2, and one subgroup of order 1. The three elements of order 2 are not central, hence the subgroups of order 2 are not normal. The commutator subgroup and the trivial subgroups are normal. The subgroup lattice of S_3 is



EXAMPLE 2.3.37. In Example 2.1.16 we defined the dihedral group D_n as the group of symmetries of a regular n-gon. For instance, if $n = 4$, the dihedral group

$$D_4 = \{e,(1234),(13)(24),(1432),(13),(24),(12)(34),(14)(23)\}$$

is a group of order 8 and is the group of symmetries of a square. In this example we use cycle notation, so $R = (1234)$ represents a rotatation of the square through an angle of 90 degrees. The horizontal flip that fixes vertex 1 is $H = (24)$. The multiplicative powers of each element of D_4 are given in the rows of the following table. The order of the element is listed in the last column.

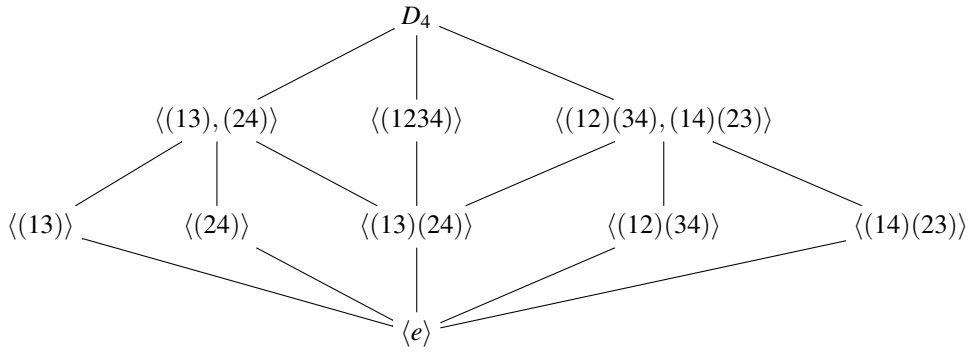
x	x^2	x^3	x^4	$\lvert x\rvert$
e				1
(1234)	$(13)(24)$	(1432)	e	4
$(13)(24)$	e			2
(1432)	$(13)(24)$	(1234)	e	4
(13)	e			2
(24)	e			2
$(12)(34)$	e			2
$(14)(23)$	e			2

There are 2 elements of order 4, 5 elements of order 2, and 1 element of order 1. Each element of order 2 generates a cyclic subgroup of order 2. The elements of order 4 are

inverses of each other and generate the only cyclic subgroup of order 4 in D_4. There are two more subgroups of order 4 that are not cyclic:

$$\langle (13),(24)\rangle = \{e,(13),(13)(24),(24)\}$$
$$\langle (12)(34),(14)(23)\rangle = \{e,(12)(34),(13)(24),(14)(23)\}.$$

The trivial subgroups $\langle e\rangle$ and D_4 are normal. The three subgroups of order 4 are normal, by Exercise 2.3.17. The center of D_4 is the cyclic subgroup $\langle (13)(24)\rangle$ and is normal, by Exercise 2.3.38. The commutator subgroup of D_4 is the cyclic subgroup $\langle (13)(24)\rangle$, by Exercise 2.3.43. The only subgroups of D_4 that are not normal are the four cyclic subgroups of order 2 that are not central. The subgroup lattice of D_4 is



where a line indicates set containment.

3.6. Exercises.

EXERCISE 2.3.38. Let G be a group. As in Definition 2.3.30, the center of G is the set $Z(G) = \{x \in G \mid xy = yx \text{ for every } y \in G\}$. Prove the following:

(1) $Z(G)$ is an abelian group.
(2) $Z(G)$ is a normal subgroup of G.
(3) If H and K are groups, then $Z(H \times K) = Z(H) \times Z(K)$.
(4) If $G/Z(G)$ is a cyclic group, then G is abelian.

EXERCISE 2.3.39. Let G be a group and $\mathrm{Aut}(G)$ the group of all automorphisms of G. As in Exercise 2.3.19, for every $a \in G$, let $\alpha_a : G \to G$ be defined by $\alpha_a(x) = a^{-1}xa$. Define $\theta : G \to \mathrm{Aut}(G)$ by $\theta(a) = \alpha_{a^{-1}}$. Show that θ is a homomorphism of groups. The image of θ is called the group of inner automorphisms of G. Show that $\ker(\theta)$ is equal to $Z(G)$, the center of G. Conclude that the group of inner automorphisms of G is isomorphic to $G/Z(G)$.

EXERCISE 2.3.40. Let $\theta : G \to G'$ be a homomorphism of groups and $x \in G$ an element of finite order. Show that $|\theta(x)|$ divides $|x|$.

EXERCISE 2.3.41. Let n be a positive integer. Prove that $\sum_{d|n} \phi(d) = n$. See Definition 1.2.15 for the notation $\sum_{d|n}$. (Hint: Apply Theorem 2.3.25.)

EXERCISE 2.3.42. Let G be a group. The *commutator subgroup of* G is the subgroup of G generated by the set $\{xyx^{-1}y^{-1} \mid x,y \in G\}$ and is denoted G'. Prove:

(1) G' is a normal subgroup of G.
(2) G/G' is abelian.
(3) If N is a normal subgroup of G such that G/N is abelian, then $G' \subseteq N$.
(4) If H is a subgroup of G and $G' \subseteq H$, then H is normal in G.

EXERCISE 2.3.43. Let $G = D_n$ be the dihedral group of order $2n$. Compute the commutator subgroup G' (see Exercise 2.3.42). (Hint: If $\sigma = (123\cdots n)$, show that G' is the cyclic group generated by σ^2.)

EXERCISE 2.3.44. Let

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 6 & 1 & 5 & 3 & 7 & 2 \end{bmatrix}, \quad \tau = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 2 & 4 & 3 & 6 & 1 & 7 \end{bmatrix}$$

be permutations in S_7. Compute $\tau\sigma\tau^{-1}$. Write σ, τ, $\tau\sigma\tau^{-1}$ using cycle notation. Show that σ factors into a 4-cycle times a 3-cycle. Show that $\tau\sigma\tau^{-1}$ factors into a 4-cycle times a 3-cycle. This is a special case of Lemma 2.6.6.

EXERCISE 2.3.45. Let G be a group and $X \subseteq G$. Let $\mathscr{S}$ be the set of all normal subgroups H in G such that $X \subseteq H$. Prove that $N = \bigcap_{H \in \mathscr{S}} H$ is a subgroup of G satisfying:

(1) N is a the smallest normal subgroup of G containing X.
(2) N is equal to the subgroup of G generated by the set $\bigcup_{g \in G} gXg^{-1}$.

We call N the *normal subgroup of G generated by X*.

EXERCISE 2.3.46. Let F be a field and $G = \mathrm{GL}_2(F)$ the general linear group of 2-by-2 matrices over F. Show that the commutator subgroup G' (see Exercise 2.3.42) is a subgroup of the special linear group $\mathrm{SL}_2(F)$ (see Example 2.3.35). For a continuation of this example, see Exercise 2.3.50.

EXERCISE 2.3.47. Let $\mathrm{GL}_2(F)$ be the general linear group of invertible 2-by-2 matrices over the field F and $\det : \mathrm{GL}_2(F) \to F^*$ the determinant function (see Example 2.1.20). Consider the following sets consisting of upper triangular matrices in $\mathrm{GL}_2(F)$:

$$U = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in M_2(F) \mid ad \neq 0 \right\},$$
$$D = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \in M_2(F) \mid b \in F \right\}.$$

(1) Show that U is a subgroup of $\mathrm{GL}_2(F)$.
(2) Show that $\det : U \to F^*$ is an epimorphism of groups and describe the kernel as a set of matrices.
(3) Show that D is isomorphic to $(F, +)$, the additive group of the field F.
(4) Show that D is a normal subgroup of U and $U/D \cong F^* \times F^*$.
(5) Show that D is equal to the commutator subgroup of U (see Exercise 2.3.42).

For a continuation of this example, see Exercise 2.3.48.

EXERCISE 2.3.48. As in Exercise 2.3.47, let F be a field, $\mathrm{GL}_2(F)$ the general linear group of 2-by-2 matrices over F, and U the subgroup of $\mathrm{GL}_2(F)$ consisting of all upper triangular invertible matrices.

(1) Define $\theta : U \to F^*$ by $\theta \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = d$. Show that θ is a group epimorphism. Let $T = \ker \theta$. Describe T as a set of matrices.
(2) Show that
$$W = \left\{ \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \in M_2(F) \mid a \in F^* \right\}$$
is a subgroup of U. Assume $F \neq \mathbb{Z}/2$. In other words, assume F contains at least three elements. Show:
   (a) W is not a normal subgroup of U.

(b) The normal subgroup of U generated by W (for this terminology, see Exercise 2.3.45) is the group T of Part (1).

For a continuation of this example, see Exercise 2.5.21.

EXERCISE 2.3.49. Let $\mathbb{C}^*$ be the group of all nonzero complex numbers under multiplication and $S^1 = \{z \in \mathbb{C} \mid |z| = 1\}$ the subgroup of all complex numbers of absolute value 1 (see Exercise 2.3.21). Show that the quotient group $\mathbb{C}^*/S^1$ is isomorphic to $(\mathbb{R}_{>0}, \cdot)$, the multiplicative abelian group of positive real numbers.

EXERCISE 2.3.50. This exercise is a continuation of Exercise 2.3.46. Let F be a field and assume $F \neq \mathbb{Z}/2$. In other words, assume F is a field that has at least three elements. Show that the commutator subgroup of $\mathrm{GL}_2(F)$, the general linear group of 2-by-2 matrices over F, is equal to $\mathrm{SL}_2(F)$, the special linear group. (Although the proof is relatively long and tedious, it is elementary and involves only material already covered in this book.)

EXERCISE 2.3.51. Let Q_8 be the quaternion 8-group of Example 2.1.18 and D_4 the dihedral group of Example 2.1.16. Let C_4 be a cyclic group of order 4. For each of the following statements, either exhibit an example to substantiate the claim, or prove that the claim is false.

(1) There exists a monomorphism of groups $C_4 \to Q_8$.
(2) There exists an epimorphism of groups $Q_8 \to C_4$.
(3) There exists a monomorphism of groups $C_4 \to D_4$.
(4) There exists an epimorphism of groups $D_4 \to C_4$.

4. Group actions

4.1. Group actions, orbits and stabilizers.

LEMMA 2.4.1. *Let G be a group and S a nonempty set. The following are equivalent.*

(1) There is a homomorphism of groups $\theta : G \to \mathrm{Perm}(S)$.
(2) There is a function $G \times S \to S$, where the image of the ordered pair (g,x) is denoted $g * x$, and the properties
(a) (associative law) $(g_1 g_2) * x = g_1 * (g_2 * x)$ for all $g_1, g_2 \in G$, $x \in S$ and
(b) ($e \in G$ acts as the identity function) $e * x = x$, for all $x \in S$
are satisfied.

PROOF. (1) implies (2): Instead of $\theta(g)(x)$ we will write $g * x$. The assignment $(g,x) \mapsto g * x$ defines a function $G \times S \to S$. Then

$$\begin{aligned}(g_1 g_2) * x &= \theta(g_1 g_2)(x) \\ &= \theta(g_1)(\theta(g_2)(x)) \\ &= g_1 * (g_2 * x)\end{aligned}$$

and $e * x = \theta(e)(x) = 1_S(x) = x$.

(2) implies (1): For each $g \in G$, define $\lambda_g : S \to S$ to be the "left multiplication by g" function defined by $\lambda_g(x) = g * x$. Since $g * g^{-1} = g^{-1} * g = e$, λ_g is a permutation of S. Define $\theta : G \to \mathrm{Perm}(S)$ by $\theta(g) = \lambda_g$. The associative law implies $\theta(g_1 g_2) = \theta(g_1)\theta(g_2)$, so θ is a homomorphism. □

In light of Lemma 2.4.1 we make the following definition.

DEFINITION 2.4.2. Let G be a group and S a nonempty set. We say *G acts on S as a group of permutations*, if there is a homomorphism of groups $\theta : G \to \mathrm{Perm}(S)$. If $g \in G$ and $x \in S$, instead of $\theta(g)(x)$ we usually write $g * x$. If θ is one-to-one, then the group action is said to be *faithful*.

EXAMPLE 2.4.3. Let G be a group. As in Example 2.1.7, if $a \in G$, then $\lambda_a : G \to G$ is the "left multiplication by a" function and λ_a is a permutation of the set G. Since $\lambda_{ab} = \lambda_a \lambda_b$, the assignment $a \mapsto \lambda_a$ defines a homomorphism of groups $\lambda : G \to \mathrm{Perm}(G)$. Proposition 2.1.6 shows that λ is one-to-one.

THEOREM 2.4.4. *(Cayley's Theorem) A finite group of order n is isomorphic to a subgroup of the symmetric group S_n.*

PROOF. Let $G = \{g_1, \ldots, g_n\}$ be a fixed enumeration of the elements of G. Then we can identify $\mathrm{Perm}(G)$ with the symmetric group S_n. By Example 2.4.3, G is isomorphic to a subgroup of S_n. □

EXAMPLE 2.4.5. Let G be a group and H a subgroup. If $xH = yH$, then $axH = ayH$ because $(ax)^{-1}ay = x^{-1}y \in H$. So $a \in G$ and $xH \in G/H$, then $a * xH = (ax)H$ defines an action by G on the set G/H by left multiplication. The reader should verify that the criteria of Lemma 2.4.1 (2) are satisfied.

LEMMA 2.4.6. *Let H and K be groups. The following are equivalent.*

(1) There is a homomorphism of groups $\theta : K \to \mathrm{Aut}(H)$.

(2) There is a function $K \times H \to H$, where the image of the ordered pair (k,x) is denoted $k * x$, and the properties

(a) (associative law) $(k_1k_2) * x = k_1 * (k_2 * x)$ for all $k_1, k_2 \in k$, $x \in H$ and

(b) ($e \in K$ acts as the identity function) $e * x = x$, for all $x \in H$

(c) (distributive law) $k * (xy) = (k * x)(k * y)$ for all $k \in K$, $x, y \in H$.

are satisfied.

PROOF. (1) implies (2): We identify $\mathrm{Aut}(H)$ with a subgroup of $\mathrm{Perm}(H)$. Then by Lemma 2.4.1, K acts on H as a group of permutations. The action by K on H is defined by $k * x = \theta(k)(x)$ and properties (a) and (b) are satisfied. The distributive law follows from the fact that $\theta(k)$ is a homomorphism if $k \in K$.

(2) implies (1): By Lemma 2.4.1, $K \to \mathrm{Perm}(H)$ is a homomorphism of groups, where $k \mapsto \lambda_k$. For $k \in K$, λ_k is a permutation of H. The distributive law implies λ_k is a homomorphism. □

In light of Lemma 2.4.6 we make the following definition.

DEFINITION 2.4.7. Let H and K be groups. We say *K acts on H as a group of automorphisms*, if there is a homomorphism of groups $\theta : K \to \mathrm{Aut}(H)$.

EXAMPLE 2.4.8. Let G be a group. If $g \in G$, then α_g is the inner automorphism of G defined by conjugation by g. That is, $\alpha_g(x) = g^{-1}xg$. By Exercise 2.3.39, there is a homomorphism of groups $G \to \mathrm{Aut}(G)$ defined by $a \mapsto \alpha_{a^{-1}}$. More generally, if N is a normal subgroup of G, and $g \in G$, then α_g restricts to an automorphism of N. Therefore there is a homomorphism $G \to \mathrm{Aut}(N)$ defined by $a \mapsto \alpha_{a^{-1}}$. See Exercise 2.4.16 for a continuation of this example.

DEFINITION 2.4.9. Let G be a group acting as a group of permutations of a nonempty set X. Define a relation $\sim$ on X by the rule $x \sim y$ if $y = g * x$ for some $g \in G$. Then $x = e * x$ implies $x \sim x$, and if $y = g * x$, then $x = g^{-1} * y$. Moreover, if $y = g_1 * x$ and $z = g_2 * y$, then

$z = g_2g_1 * x$. This proves that $\sim$ is an equivalence relation on X. The equivalence class of x is called the *orbit of* x. The orbit of x is equal to $G*x = \{g*x \mid g \in G\}$. The set of orbits is denoted X/G. If $x \in X$, then the *stabilizer of* x *in* G is $G_x = \{g \in G \mid g*x = x\}$. By Theorem 2.4.10, G_x is a subgroup of G, therefore, G_x is sometimes called the *subgroup fixing* x. If $G_x = G$, then we say x *is fixed by* G. The set $X_0 = \{x \in X \mid g*x = x \text{ for all } g \in G\}$ is the set of all x in X that are fixed by G.

THEOREM 2.4.10. *Let G be a group acting on a nonempty set X. If $x \in X$, then G_x, the stabilizer of x in G satisfies the following properties.*

(1) G_x is a subgroup of G.
(2) The length of the orbit $G*x$ is equal to the index $[G : G_x]$.

PROOF. (1): Since $e \in G_x$, we have $G_x \neq \emptyset$. If $a,b \in G_x$, then $ab*x = a*(b*x) = a*x = x$, hence $ab \in G_x$. If $a*x = x$, then $x = a^{-1}*x$. This proves G_x is a subgroup of G.

(2): We show that there is a one-to-one correspondence between the set of left cosets of G_x in G and the set $G*x$. Define a function $f : G \to G*x$ by $f(g) = g*x$. Then f is onto. Define a relation on G by the rule: $g \approx h$ if and only if $f(g) = f(h)$. By Exercise 1.1.14, $\approx$ is an equivalence relation. Notice that $g \approx h$ if and only if $g^{-1}h \in G_x$, which is equivalent to $g \equiv h \pmod{G_x}$. Therefore, $\bar{f} : G/G_x \to G*x$ is a one-to-one correspondence. □

4.2. Conjugates and the Class Equation.

EXAMPLE 2.4.11. Let G be a group and $X = 2^G$ the power set of G. If S is a subset of G, and $a \in G$, then $a*S = aSa^{-1}$ defines an action by G on X. The associative law is $ab*S = abS(ab)^{-1} = a(bSb^{-1})a^{-1} = a*(b*S)$. The stabilizer of S in G is usually called the *normalizer of* S *in* G and is denoted $N_G(S) = \{a \in G \mid aSa^{-1} = S\}$. The orbit of S under this action is the set $\{a^{-1}Sa \mid a \in G\}$ of all distinct conjugates of S by elements of G.

PROPOSITION 2.4.12. *Let G be a group and S a subset of G. The normalizer of S in G satisfies the following properties.*

(1) $N_G(S)$ is a subgroup of G.
(2) If H is a subgroup of G, then $N_G(H)$ is the largest subgroup of G containing H as a normal subgroup.
(3) The number of distinct conjugates of S by elements in G is $[G : N_G(S)]$.

PROOF. (1) and (3): These follow from Theorem 2.4.10.

(2): Since H is a subgroup, $a^{-1}Ha = H$ for all $a \in H$. Therefore, $H \subseteq N_G(H)$. If $x \in N_G(H)$, then $x^{-1}Hx = H$. Therefore, H is normal in $N_G(H)$. Suppose $H \leq K \leq G$ and H is a normal subgroup of K. For all $x \in K$, $x^{-1}Hx = H$, hence $K \subseteq N_G(H)$. □

Let G be a group acting on itself by conjugation. If $x \in G$, the orbit of x is $C_x = \{a^{-1}xa \mid a \in G\}$ and is called the *conjugacy class of* x. The number of conjugates of x is the length of the orbit C_x. By Theorem 2.4.10, $|C_x| = [G : N_G(x)]$. If x is in $Z(G)$, the center of G, then $N_G(x) = G$ and $C_x = \{x\}$. Since $|G|$ is finite, there are a finite number of conjugacy classes. If $x_1, \ldots, x_n$ is a full set of representatives for the conjugacy classes that are not in $Z(G)$, then $G = Z(G) \cup (G - Z(G)) = Z(G) \cup \left(\cup_{i=1}^{n} C_{x_i}\right)$ is a disjoint union. Taking cardinalities of both sides of this equation yields the next corollary.

COROLLARY 2.4.13. *(The Class Equation) Let G be a finite group and $x_1, \ldots, x_n$ a full set of representatives for the conjugacy classes that are not in $Z(G)$. Then*

$$|G| = |Z(G)| + \sum_{i=1}^{n} [G : N_G(x_i)].$$

As an application of Corollary 2.4.13, we prove Cauchy's Theorem. Recall that we already proved Theorem 2.3.28, which is the abelian version of this result. A second more concise proof of Cauchy's Theorem is given below in Theorem 2.7.3.

COROLLARY 2.4.14. *(Cauchy's Theorem) Let G be a finite group of order n and p a prime divisor of n. Then G contains an element of order p.*

PROOF. The proof is by induction on n. If G is abelian, then G has an element of order p, by Theorem 2.3.28. Inductively assume $n \geq 6$, G is nonabelian, and that the result holds for any group of order less than n. Let $x_1, \ldots, x_m$ be a full set of representatives for the conjugacy classes that are not in $Z(G)$. By our induction hypothesis, $m \geq 1$. Solving the Class Equation of Corollary 2.4.13 for $|Z(G)|$, we have

$$|Z(G)| = |G| - \sum_{i=1}^{m} [G : N_G(x_i)]. \tag{4.1}$$

For each x_i, $N_G(x_i)$ is a proper subgroup of G. If p divides $|N_G(x_i)|$ for some i, then by our induction hypothesis, there is an element of order p in $N_G(x_i)$. Therefore, assume for every i that p does not divide $|N_G(x_i)|$. By Corollary 2.2.12, $|G| = |N_G(x_i)|[G : N_G(x_i)]$. Since p divides $|G|$ and p does not divide $|N_G(x_i)|$, we have p divides $[G : N_G(x_i)]$, for every i. Therefore, p divides the right hand side of (4.1). Hence p divides $|Z(G)|$. By Theorem 2.3.28, we know that $Z(G)$ has an element of order p. □

4.3. Exercises.

EXERCISE 2.4.15. Let H and K be groups. Recall (Definition 2.4.7) that we say K acts as a group of automorphisms of H if there is a homomorphism of groups $\theta : K \to \mathrm{Aut}(H)$. In this case, write $k * x$ instead of $\theta(k)(x)$. Prove the following:

(1) $k * e = e$ for all $k \in K$.
(2) $(k * x)^{-1} = k * x^{-1}$ for all $k \in K$, $x \in H$.

EXERCISE 2.4.16. Let G be a group containing a normal subgroup N. Let K be an arbitrary subgroup of G. Generalize Example 2.4.8 by showing that K acts on N as a group of automorphisms. Specifically, show that if $k \in K$ and $x \in N$, then $k * x = kxk^{-1}$ defines an action by K on N as a group of automorphisms.

EXERCISE 2.4.17. (Semidirect product) As in Definition 2.4.7, let H and K be groups and assume K acts on H as a group of automorphisms. Define a binary operation on $H \times K$ by the rule:

$$(x_1, k_1)(x_2, k_2) = (x_1(k_1 * x_2), k_1 k_2).$$

(1) Show that the binary operation defined above makes $H \times K$ into a group where the identity element is (e, e) and the inverse of (x, k) is $(k^{-1} * x^{-1}, k^{-1})$. This group is denoted $H \rtimes K$ and is called the *semidirect product* of H and K.
(2) Show that $N = \{(x, e) \mid x \in H\}$ is a normal subgroup of $H \rtimes K$ and the quotient $(H \rtimes K)/N$ is isomorphic to K. Show that H is isomorphic to N.
(3) Show that $C = \{(e, k) \mid k \in K\}$ is a subgroup of $H \rtimes K$ and K is isomorphic to C.

EXERCISE 2.4.18. Let G be a group containing subgroups N and K satisfying:

(1) $G = NK$,
(2) N is normal in G, and
(3) $N \cap K = \langle e \rangle$.

As in Exercise 2.4.16, let K act on N by conjugation. Prove that the semidirect product $N \rtimes K$ (see Exercise 2.4.17) is isomorphic to G.

EXERCISE 2.4.19. Let $Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$ be the quaternion 8-group of Example 2.1.18. Show that every subgroup of Q_8 is normal. Let Z denote the center of Q_8. Show that Z is a group of order two and is contained in every nontrivial subgroup of Q_8. Show that Q_8 is not a semidirect product of two subgroups.

EXERCISE 2.4.20. Let $m,n \in \mathbb{N}$ be positive integers. Show that there are $\gcd(m,n)$ distinct homomorphisms from $\mathbb{Z}/m$ to $\mathbb{Z}/n$. See Exercises 3.1.17 and 2.8.12 for a continuation of this exercise.

EXERCISE 2.4.21. If $n \geq 3$, show that the dihedral group D_n is isomorphic to the semidirect product of a cyclic subgroup of order n and a cyclic subgroup of order two.

EXERCISE 2.4.22. Let p be an odd prime. Let G be a group of order $2p$. Show that G has a unique subgroup of order p. Denote by P the subgroup of G of order p. Show that G is isomorphic to the semidirect product of P and a cyclic subgroup of order two that acts on P by conjugation. Show that G is isomorphic to either the cyclic group $\mathbb{Z}/2p$ or the dihedral group D_p.

EXERCISE 2.4.23. Show how to construct a nonabelian group of order $9 \cdot 37$ that contains a cyclic subgroup of order 9 and a cyclic subgroup of order 37.

EXERCISE 2.4.24. Let G be a group acting on a set X (see Definition 2.4.2). Let $G_0 = \{g \in G \mid g * x = x \text{ for all } x \in X\}$. Show that G_0 is a normal subgroup of G.

EXERCISE 2.4.25. Let G be a group and H a subgroup of G. As in Example 2.4.5, G acts on G/H by left multiplications. By Lemma 2.4.1, there is a a homomorphism of groups $\theta : G \to \mathrm{Perm}(G/H)$. As in Exercise 2.4.24, denote the kernel of θ by G_0. Show that G_0 is a normal subgroup of G contained in H.

EXERCISE 2.4.26. Let p be a prime and G be a group of order p^2. Apply Exercise 2.4.25 to show that every subgroup of G is normal. If G has order p^r, $r > 1$, show that every subgroup of order p^{r-1} is normal in G.

EXERCISE 2.4.27. Let p and q be primes such that $q \equiv 1 \pmod{p}$. Show how to construct a nonabelian group of order pq.

EXERCISE 2.4.28. Let $Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$ be the quaternion 8-group of Example 2.1.18. Show that $Q_8 = \{1\} \cup \{-1\} \cup \{\pm i\} \cup \{\pm j\} \cup \{\pm k\}$ is the decomposition of Q_8 into conjugacy classes.

EXERCISE 2.4.29. The group of symmetries of a square is

$$D_4 = \{e, (1234), (13)(24), (1432), (12)(34), (14)(23), (13), (24)\}.$$

Show that $D_4 = \{e\} \cup \{(13)(24)\} \cup \{(1234), (1432)\} \cup \{(24), (13)\} \cup \{(12)(34), (14)(23)\}$ is the decomposition of D_4 into conjugacy classes.

EXERCISE 2.4.30. The group of symmetries of a regular pentagon is

$$\begin{aligned} D_5 = \{&e, (12345), (13524), (14253), (15432), \\ &(25)(34), (15)(24), (13)(45), (12)(35), (14)(23)\}. \end{aligned}$$

Show that

$$\begin{aligned} D_5 = \{e\} &\cup \{(12345), (15432)\} \cup \{(13524), (14253)\} \\ &\cup \{(25)(34), (15)(24), (13)(45), (12)(35), (14)(23)\} \end{aligned}$$

is the decomposition of D_5 into conjugacy classes.

EXERCISE 2.4.31. Show how to construct two nonisomorphic nonabelian groups of order 40 each of which is a semidirect product of two cyclic groups.

EXERCISE 2.4.32. Let G be a finite group and H a subgroup of G. Suppose the only normal subgroup of G contained in H is $\langle e\rangle$. Show that G is isomorphic to a subgroup of S_n, where $n = [G : H]$. (Hint: Exercise 2.4.25.)

EXERCISE 2.4.33. For the following choices of p and q, show how to construct a nonabelian group of order pq which is a semidirect product of two cyclic groups.

(1) $p = 5, q = 11$.
(2) $p = 7, q = 29$.

EXERCISE 2.4.34. Let p be a prime number and n an integer such that $0 < n < p$. If G is a finite group of order pn and P is a subgroup of order p, then P is normal. (Hint: Exercise 2.4.25.)

5. Direct products

5.1. External direct product.

DEFINITION 2.5.1. Let I be an index set and $\{G_i \mid i \in I\}$ a family of multiplicative groups indexed by I. Although the groups G_i in general are not equal as sets and have no common elements, we abuse notation and use the same symbol e to denote the identity element of each group G_i. The cartesian product is $\prod_{i\in I} G_i = \{f : I \to \cup_{i\in I} \mid f(i) \in G_i\}$. The cartesian product is a group if the binary operation is defined to be coordinate-wise multiplication: $(fg)(i) = f(i)g(i)$. The identity element is the constant function $e(i) = e$ and the inverse of f is defined by $f^{-1}(i) = (f(i))^{-1}$, the coordinate-wise inverse. The group $\prod_{i\in I} G_i$ is called the *direct product*. Sometimes $\prod_{i\in I} G_i$ is called the *external direct product* to distinguish it from the construction in Definition 2.5.3 below. For every $k \in I$ there is a *canonical injection map* $\iota_k : G_k \to \prod_{i\in I} G_i$ which maps $x \in G_k$ to $\iota_k(x)$, where

$$\iota_k(x)(i) = \begin{cases} x & \text{if } i = k \\ e & \text{otherwise.} \end{cases}$$

The *canonical projection map* is $\pi_k : \prod_{i\in I} G_i \to G_k$ where $\pi_k(f) = f(k)$. The reader should verify that ι_k is a monomorphism, π_k is an epimorphism and $\pi_k\iota_k = 1_{G_k}$.

When $I = \{1,\dots,n\}$ is a finite set, the direct product is identified with the set of n-tuples $\{(x_1,\dots,x_n) \mid x_i \in G_i\}$ and it is written $G_1 \times \cdots \times G_n$ or $\prod_{i=1}^n G_i$. Multiplication is defined coordinate-wise, hence $(x_1,\dots,x_n)(y_1,\dots,y_n) = (x_1y_1,\dots,x_ny_n)$. The identity element is $(e,\dots,e)$, and $(x_1,\dots,x_n)^{-1}$ is $(x_1^{-1},\dots,x_n^{-1})$.

THEOREM 2.5.2. *(Chinese Remainder Theorem) Let m and n be positive integers and let*

$$\psi : \mathbb{Z} \to \mathbb{Z}/m \times \mathbb{Z}/n$$

be defined by $\psi(x) = (\eta_m(x), \eta_n(x))$, *where* $\eta_m : \mathbb{Z} \to \mathbb{Z}/m$ *and* $\eta_n : \mathbb{Z} \to \mathbb{Z}/n$ *are the natural maps. Then the following are true:*

(1) $\ker(\psi) = \langle M\rangle$, *where* $M = \operatorname{lcm}(m,n)$.
(2) ψ *is onto if and only if* $\gcd(m,n) = 1$.
(3) $\mathbb{Z}/m \times \mathbb{Z}/n$ *is cyclic if and only if* $\gcd(m,n) = 1$.

PROOF. (1): Since η_m and η_n are homomorphisms, it is routine to verify that ψ is a homomorphism. By Theorem 2.2.15, the kernel of η_m is $m\mathbb{Z}$ and the kernel of η_n is

$n\mathbb{Z}$. We see that $\ker(\psi) = \ker(\eta_m) \cap \ker(\eta_n)$ is equal to $\{x \in \mathbb{Z} \mid m \mid x \text{ and } n \mid x\}$. By Theorem 2.2.15, $\ker(\psi)$ is generated by $M = \operatorname{lcm}(m,n)$.

(2): Let $d = \gcd(m,n)$. By Proposition 1.2.10, $Md = mn$. By Theorem 2.3.12, $\operatorname{im}(\psi)$ is isomorphic to $\mathbb{Z}/M$, which has order M. We see that ψ is onto if and only if $M = mn$, which is true if and only if $d = 1$.

(3): If $d = 1$, then the direct product $\mathbb{Z}/m \times \mathbb{Z}/n$ is cyclic by (2). Assume $d > 1$. To show the direct product is not cyclic, we show that it contains more than $\phi(d)$ elements of order d and apply Theorem 2.3.25 (6). Let $A = \{x \in \mathbb{Z}/m \mid |x| = d\}$. Then $|A| = \phi(d)$. If $x \in A$, then by an application of Lemma 2.2.16 (5) we see that $(x,0)$ has order d in the direct product. Likewise, if $B = \{y \in \mathbb{Z}/n \mid |y| = d\}$, then $|B| = \phi(d)$ and $(0,y)$ has order d, for each $y \in B$. Therefore, the direct product contains at least $2\phi(d)$ elements of order d. This proves (3). □

5.2. Internal direct product.

DEFINITION 2.5.3. Let G be a group and $N_1, N_2, \ldots, N_m$ a collection of subgroups of G satisfying:

(1) N_i is a normal subgroup of G for each i,
(2) $G = N_1 N_2 \cdots N_m$, and
(3) if $x_i \in N_i$ for each i and $e = x_1 x_2 \cdots x_m$, then $x_i = e$ for each i.

Then we say G is the *internal direct product* of $N_1, \ldots, N_m$.

LEMMA 2.5.4. *Suppose G is the internal direct product of $N_1, N_2, \ldots, N_m$. Then the following are true.*

(1) If $i \neq j$, then $N_i \cap N_j = \langle e \rangle$.
(2) If $i \neq j$, $x_i \in N_i$, $x_j \in N_j$, then $x_i x_j = x_j x_i$.
(3) For each i let $x_i, y_i \in N_i$. If $x = x_1 x_2 \cdots x_m$, and $y = y_1 y_2 \cdots y_m$, then
  (a) $xy = (x_1 y_1)(x_2 y_2) \cdots (x_m y_m)$, and
  (b) $x^{-1} = x_1^{-1} x_2^{-1} \cdots x_m^{-1}$.
(4) If $x \in G$, then x has a unique representation as a product $x = x_1 x_2 \cdots x_m$, where $x_i \in N_i$ for each i.
(5) G is isomorphic to the (external) direct product $N_1 \times N_2 \times \cdots \times N_m$.

PROOF. (1): Let $x \in N_i \cap N_j$. Assume $1 \leq i < j \leq m$. In the product $N_1 \cdots N_i \cdots N_j \cdots N_m$ we have

$$e = e \cdots x \cdots x^{-1} \cdots e$$

where the ith factor is x, the jth factor is x^{-1}, and all other factors are the group identity e. By the uniqueness property of Definition 2.5.3, $x = e$.

(2): Because N_i and N_j are normal in G, we have $x_i y_j x_i^{-1} x_j^{-1}$ is in $N_i \cap N_j = \langle e \rangle$.

(3): The two identities follow immediately from Part (2).

(4): Assume $x = x_1 x_2 \cdots x_m$, where $x_i \in N_i$ for each i. Assume $x = y_1 y_2 \cdots y_m$, where $y_i \in N_i$ for each i is another such representation. Using Part (3), we get

$$e = x x^{-1} = (x_1 y_1^{-1}) \cdots (x_m y_m^{-1}).$$

By the uniqueness property of Definition 2.5.3, $x_i = y_i$ for each i.

(5): Let $\psi : N_1 \times N_2 \times \cdots \times N_m \to G$ be the function defined by multiplication in the group G: $\psi(x_1, x_2, \ldots, x_m) = x_1 x_2 \cdots x_m$. By Part (3), ψ is a homomorphism. By Definition 2.5.3, ψ is a one-to-one correspondence. □

PROPOSITION 2.5.5. *Let G be a group and $N_1, \ldots, N_m$ a collection of normal subgroups. Then the following are equivalent.*

(1) G is the internal direct product of $N_1, \ldots, N_m$.
(2) The function $\phi : N_1 \times \cdots \times N_m \to G$ defined by $\phi(x_1, \ldots, x_m) = x_1 \cdots x_m$ is an isomorphism of groups.
(3) $G = N_1 \cdots N_m$ and for each k, the intersection $N_k \cap (N_1 \cdots N_{k-1} N_{k+1} \cdots N_m)$ is the trivial subgroup $\langle e \rangle$.
(4) $G = N_1 \cdots N_m$, and $N_1 \cap N_2 \cdots N_m = N_2 \cap N_3 \cdots N_m = \cdots = N_{m-1} \cap N_m = \langle e \rangle$.

PROOF. (1) implies (2): This is Lemma 2.5.4 (5).

(2) implies (3): Since ϕ is onto we have $G = N_1 \cdots N_m$. Let x be an arbitrary element of $N_k \cap (N_1 \cdots N_{k-1} N_{k+1} \cdots N_m)$. We can write x in two ways: $x = x_k \in N_k$, and $x = x_1 \cdots x_{k-1} x_{k+1} \cdots x_m \in N_1 \cdots N_{k-1} N_{k+1} \cdots N_m$. Therefore $x = \phi(e, \ldots, e, x_k, e, \ldots, e) = \phi(x_1, \ldots, x_{k-1}, e, x_{k+1}, \ldots, x_m)$. Since ϕ is one-to-one, $x = x_k = e$.

(3) implies (4): For each $k = 1, \ldots, m-1$ we have: $N_{k+1} \cdots N_m \subseteq N_1 \cdots N_{k-1} N_{k+1} \cdots N_m$. Therefore, $N_k \cap (N_{k+1} \cdots N_m) \subseteq N_k \cap (N_1 \cdots N_{k-1} N_{k+1} \cdots N_m) = \langle e \rangle$.

(4) implies (1): Let $e = x_1 x_2 \cdots x_m$ be a representation of e in $N_1 N_2 \cdots N_m$. Then $x_1^{-1} = x_2 \cdots x_m$ is in $N_1 \cap N_2 \cdots N_m = \langle e \rangle$. Therefore, $x_1 = e$ and $x_2 \cdots x_m = e$. Inductively, assume $1 < k < m$ and $x_k \cdots x_m = e$. Then $x_k^{-1} = x_{k+1} \cdots x_m$ is in $N_k \cap N_{k+1} \cdots N_m = \langle e \rangle$. Therefore, $x_k = e$ and $x_{k+1} \cdots x_m = e$. By induction, we are done. □

5.3. Free Groups. Let X be a set, which will be called the *alphabet*. A *word* on the alphabet X is a finite string of the form

$$w = a_1^{\varepsilon_1} a_2^{\varepsilon_2} \cdots a_n^{\varepsilon_n}$$

where $n \geq 0$, each a_i is an element of X and $\varepsilon_i \in \{-1, 1\}$. The *length* of the string is n. The only string of length 0 is called the *empty string* and is denoted e. A string is *reduced* if it contains no substrings of the form xx^{-1} or $x^{-1}x$, for $x \in X$. Every word can be reduced in a unique way by recursively striking out all of the substrings of the form xx^{-1} or $x^{-1}x$.

LEMMA 2.5.6. *Let $v = a_1^{\varepsilon_1} a_2^{\varepsilon_2} \cdots a_n^{\varepsilon_n}$ and $w = b_1^{\phi_1} b_2^{\phi_2} \cdots b_p^{\phi_p}$ be reduced words on the alphabet X. There exist factorizations of v and w into substrings $v = v_1 v_2$, $w = w_1 w_2$ such that $v_2 w_1$ reduces to the empty word e and the reduction of vw is equal to $v_1 w_2$. The factors v_1, v_2, w_1, w_2 are unique.*

PROOF. If v has length $n = 0$, then take $v_1 = v_2 = w_1 = e$ and $w_2 = w$. In this case, $vw = v_1 w_2$ and we are done. Inductively assume $n > 0$ and that the result holds for any reduced word of length $n-1$. If $a_n^{\varepsilon} \neq b_1^{-\phi_1}$, then vw is reduced. In this case, take $v = v_1$,$v_2 = w_1 = e$, and $w_2 = w$. Otherwise, delete a_n^{ε} from the end of v and $b_1^{-\phi_1}$ from the front of w, and apply the induction hypothesis to obtain factorizations:

$$a_1^{\varepsilon_1} a_2^{\varepsilon_2} \cdots a_{n-1}^{\varepsilon_{n-1}} = v_1 v_3$$
$$b_2^{\phi_2} \cdots b_p^{\phi_p} = w_3 w_2$$

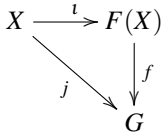
Setting $v_2 = v_3 a_n^{\varepsilon}$ and $w_1 = b_1^{\phi_1} w_3$, we have $v_2 w_1 = v_3 a_n^{\varepsilon} b_1^{\phi_1} w_3$ reduces to $v_3 w_3$ which reduces to the empty word e. Also, the reduction of vw is equal to the reduction of $v_1 v_3 w_3 w_2$ which is equal to $v_1 w_2$. This proves the existence of the factorization. The uniqueness of v_3 and w_3 implies the uniqueness of v_2 and w_1. □

LEMMA 2.5.7. *Let $F(X)$ be the set of all reduced words on X. Then $F(X)$ is a group, where the product of two words is the word defined by juxtaposition followed by reduction. The identity element for the group $F(X)$ is the empty string e. The inverse of the string*

$a_1^{\varepsilon_1}a_2^{\varepsilon_2}\cdots a_n^{\varepsilon_n}$ *is the string* $a_n^{-\varepsilon_n}\cdots a_2^{-\varepsilon_2}a_1^{-\varepsilon_1}$. *We call* $F(X)$ *the* free group on the set X. *There is a natural injection* $\iota : X \to F(X)$ *defined by* $\iota(x) = x$.

PROOF. By Lemma 2.5.6, if v and w are reduced words in $F(X)$, then the reduction of the word vw is uniquely defined. Since this binary operation does not depend on grouping by parentheses, it is associative. The rest is left to the reader. □

THEOREM 2.5.8. *(Universal Mapping Property) Let* X *be a set and* $\iota : X \to F(X)$ *the natural injection map. For any group* G *and any function* $j : X \to G$, *there is a unique homomorphism* $f : F(X) \to G$ *such that the diagram*



commutes.

PROOF. Let $v = a_1^{\varepsilon_1}a_2^{\varepsilon_2}\cdots a_n^{\varepsilon_n}$ be a reduced word in $F(X)$. Then we define $f(v)$ to be $j(a_1)^{\varepsilon_1}j(a_2)^{\varepsilon_2}\cdots j(a_n)^{\varepsilon_n}$. Then f is a well defined function and $f\iota = j$. To see that f is a homomorphism of groups, let $w = b_1^{\phi_1}b_2^{\phi_2}\cdots b_p^{\phi_p}$ be another reduced word on the alphabet X. As in Lemma 2.5.6, factor $v = v_1v_2$, $w = w_1w_2$ such that the reduction of vw is equal to v_1w_2. Since $f(v) = f(v_1v_2) = f(v_1)f(v_2)$, $f(w) = f(w_1w_2) = f(w_1)f(w_2)$, and $f(v_2)f(w_1) = e$, it follows that

$$f(vw) = f(v_1w_2) = f(v_1)f(w_2) = f(v_1)f(v_2)f(w_1)f(w_2) = f(v)f(w).$$

To prove the uniqueness claim, assume $g : F(X) \to G$ is another homomorphism and $g\iota = j$. Then $f(x) = g(x)$ for every $x \in X$. Since X is a generating set for the group $F(X)$, f is equal to g. □

COROLLARY 2.5.9. *Every group* G *is the homomorphic image of a free group.*

PROOF. In Theorem 2.5.8, take $X = G$ and $j : G \to G$ the identity map. Since j is onto, f is onto. □

DEFINITION 2.5.10. Let X be a set and Y a subset of $F(X)$. As in Exercise 2.3.45, let N be the normal subgroup of $F(X)$ generated by Y. Consider the quotient group $G = F(X)/N$. We say G is *defined by the generators* X *subject to the relations* Y. We denote the group $G = F(X)/N$ by $\langle X \mid Y\rangle$.

EXAMPLE 2.5.11. In the notation of Theorem 2.3.25, let $A = \langle a\rangle$ be a cyclic group. If A is infinite, then a presentation of A in terms of generators and relations is $A = \langle a \mid \emptyset\rangle$. If A has order $n > 0$, then a presentation of A in terms of generators and relations is $A = \langle a \mid a^n\rangle$. It is common for the relations to be written as equations. Then $A = \langle a \mid a^n = e\rangle$.

EXAMPLE 2.5.12. Let $n > 2$ and D_n the dihedral group of order $2n$ of Example 2.1.16. Then D_n is generated by two elements, R and H. The order of R is n and the order of H is 2. The so-called commutator identity is $HRH = R^{-1}$. Therefore,

$$D_n = \langle R, H \mid H^2 = e, R^n = e, HRH = R^{-1}\rangle$$

is a presentation of D_n in terms of generators and relations.

EXAMPLE 2.5.13. Let V be the Klein 4-group of Example 2.1.21. Then V is an abelian group of order 4, generated by two elements of order two. Hence,

$$V = \langle a, b \mid a^2 = b^2 = e, ab = ba \rangle$$

is a presentation in terms of generators and relations.

EXAMPLE 2.5.14. Let $Q_8 = \{1, -1, i, -i, j, -j, k, -k\}$ be the quaternion eight group of Example 2.1.18. The multiplication rules are: $(-1)^2 = 1$, $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$. So we see that Q_8 is generated by i and j. Both i and j have order 4 and $-1 = i^2 = j^2$. The commutator relation for i and j is $ij = -ji = j^3 i$. If we write a and b instead of i and j, then a presentation in terms of generators and relations is

$$Q_8 = \langle a, b \mid a^4 = e, b^4 = e, a^2 = b^2, ab = b^3 a \rangle.$$

5.4. Exercises.

EXERCISE 2.5.15. The general linear group of 2-by-2 matrices over the field $\mathbb{Z}/3$, denoted $\mathrm{GL}_2(\mathbb{Z}/3)$, is the multiplicative group of invertible matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with entries in the field $\mathbb{Z}/3$. Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, $C = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}$, $P = \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix}$, and $Q = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ be matrices with entries in $\mathbb{Z}/3$. For the following computations, access to a computer algebra system such as [**14**] is not required, but will be beneficial, especially for parts (6) and (7).

(1) Show that A, B, C, P, and Q are in $\mathrm{GL}_2(\mathbb{Z}/3)$.
(2) Compute the cyclic subgroups $\langle A \rangle$, $\langle B \rangle$, $\langle C \rangle$, $\langle P \rangle$, $\langle Q \rangle$.
(3) Show that P is in the normalizer of $\langle A \rangle$. Show that P and A generate a subgroup of order 16.
(4) Show that P is in the normalizer of $\langle B \rangle$. Show that P and B generate a subgroup of order 16.
(5) Show that Q is in the normalizer of $\langle C \rangle$. Show that Q and C generate a subgroup of order 16.
(6) If $G = \mathrm{GL}_2(\mathbb{Z}/3)$, show that G has order 48. Show that G has 3 subgroups of order 16. Show that G has 4 subgroups of order 3.
(7) The special linear group of 2-by-2 matrices over $\mathbb{Z}/3$, denoted $\mathrm{SL}_2(\mathbb{Z}/3)$, is the subgroup of $\mathrm{GL}_2(\mathbb{Z}/3)$ consisting of those matrices with determinate equal to 1. Let $S = \mathrm{SL}_2(\mathbb{Z}/3)$. Show that S has order 24. Show that S has 3 subgroups of order 8. Show that every subgroup of order 8 is isomorphic to the quaternion 8-group, $Q_8 = \{\pm 1, \pm i, \pm j, \pm k\}$ of Example 2.1.18. . Show that S has 4 subgroups of order 3.

EXERCISE 2.5.16. Give an example of a group G and subgroups $N_1, N_2, \ldots, N_m$ of G satisfying:

(1) N_i is a normal subgroup of G for each i,
(2) $G = N_1 N_2 \cdots N_m$, and
(3) if $i \neq j$, then $N_i \cap N_j = \langle e \rangle$,

such that G is not the internal direct product of $N_1, N_2, \ldots, N_m$.

EXERCISE 2.5.17. Let G be a finite abelian group. Assume G is the internal direct product of cyclic subgroups $A = \langle a \rangle$ and $B = \langle b \rangle$ where a and b both have order 6.

(1) Show that $|G| = 36$.
(2) Show that $C = \langle ab^2 \rangle$ has order 6.
(3) Compute $|AC|$.
(4) Show that $|AC|$ is the internal direct product of A and $\langle b^2 \rangle$.

EXERCISE 2.5.18. Let A and B be normal subgroups of G such that $G = AB$. Prove that $G/(A \cap B)$ is isomorphic to $G/A \times G/B$.

EXERCISE 2.5.19. Let G be a group containing subgroups A and B such that

(1) $G = AB$,
(2) $xy = yx$ for every $x \in A$ and $y \in B$, and
(3) $A \cap B = \langle e \rangle$.

Show that G is the internal direct product of A and B.

EXERCISE 2.5.20. Let A and B be groups. Let A_0 be a normal subgroup of A and B_0 a normal subgroup of B. Show that there is an isomorphism of groups

$$\frac{A \times B}{A_0 \times B_0} \cong \frac{A}{A_0} \times \frac{B}{B_0}.$$

EXERCISE 2.5.21. This is a continuation of Exercise 2.3.48. Let F be a field and

$$U = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in M_2(F) \mid ad \neq 0 \right\}$$

the set of all upper triangular matrices in $\mathrm{GL}_2(F)$. Let T be the kernel of the homomorphism $U \to F^*$ defined by $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \mapsto d$. As in Example 2.3.34, let $\delta : F^* \to \mathrm{GL}_2(F)$ be the diagonal map. Let $Z = \operatorname{im} \delta$. Show that U is the internal direct product of T and Z.

6. Permutation Groups

The group of all permutations of $\mathbb{N}_n = \{1, 2, 3, \ldots, n\}$ is called the symmetric group on n letters and is denoted S_n (see Example 2.1.14).

6.1. The cycle decomposition of a permutation. Let $\alpha = (a_1, \ldots, a_s)$ be an s-cycle and $\beta = (b_1, \ldots, b_t)$ a t-cycle. We say α and β are *disjoint* if $\{a_1, \ldots, a_s\} \cap \{b_1, \ldots, b_t\} = \emptyset$. If this is the case, then $\beta(a_i) = a_i$ for each i, and $\alpha(b_j) = b_j$ for each j. Therefore, $\alpha\beta = \beta\alpha$. This proves Lemma 2.6.1.

LEMMA 2.6.1. *If α and β are disjoint cycles in S_n, then α and β commute. That is, $\alpha\beta = \beta\alpha$.*

EXAMPLE 2.6.2. Here is an example with $n = 6$. In S_6, let

$$\alpha = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 1 & 2 & 6 & 5 \end{bmatrix}, \quad \beta = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 5 & 4 & 3 & 1 & 2 \end{bmatrix}.$$

Then $A = \langle \alpha \rangle$ acts on $\{1,2,3,4,5,6\}$. Given $x \in \{1,2,3,4,5,6\}$, the orbit of x is $A * x$. We compute the orbit decomposition under this action. The reader should verify that $A * 1 = \{1,3\}$, $A * 2 = \{2,4\}$, $A * 5 = \{5,6\}$. In Theorem 2.6.3 we find that from the orbit decomposition we can construct the factorization of α into cycles. For instance, $\alpha = (1,3)(2,4)(5,6)$. Likewise, for $B = \langle \beta \rangle$, we find the disjoint orbits are $B * 1 = \{1,6,2,5\}$, $B * 3 = \{3,4\}$ and the factorization of β into cycles is $\beta = (1,6,2,5)(3,4)$.

THEOREM 2.6.3. *If $\sigma \in S_n$ is a permutation on n letters, then σ can be written as the product of disjoint cycles. This representation is unique in the sense that if $\sigma \neq e$ and $\sigma = \alpha_1\alpha_2\cdots\alpha_k$ is a product of disjoint cycles all of length two or more and $\sigma = \beta_1\beta_2\cdots\beta_\ell$ is another such representation, then $k = \ell$ and $\beta_1, \beta_2, \ldots, \beta_k$ can be relabeled such that $\alpha_i = \beta_i$ for each i.*

PROOF. Let $\sigma \in S_n$ and let $S = \langle\sigma\rangle$. Then S acts on $\mathbb{N}_n = \{1, 2, \ldots, n\}$. Let a be an arbitrary element of $\mathbb{N}_n$. We associate to the orbit of a under S a cyclic permutation α_a. Let S_a be the subgroup of S fixing a. Then S_a is a cyclic subgroup of S. If $[S : S_a] = w$, then by Theorem 2.3.25, S_a is the unique subgroup of S with index w and $S_a = \langle\sigma^w\rangle$. By Theorem 2.4.10, the length of the orbit of a is equal to w and the orbit of a is $\{a, \sigma(a), \sigma^2(a), \ldots, \sigma^{w-1}(a)\}$. On this set σ is equal to the cyclic permutation $\alpha_a = (a, \sigma(a), \sigma^2(a), \ldots, \sigma^{w-1}(a))$. We see that for every orbit under the S-action there is an associated cyclic permutation. If $\{a_1, a_2, \ldots, a_k\}$ is a full set of representatives for the orbits, then σ is equal to the product of cycles $\alpha_{a_1}\alpha_{a_1}\cdots\alpha_{a_k}$. The orbits are disjoint, hence so are the cycles in this factorization. The uniqueness claim follows from the fact that the cycle decomposition is determined by the orbit decomposition which is uniquely determined by σ. □

COROLLARY 2.6.4. *If $\alpha_1, \alpha_2, \ldots, \alpha_m$ are pairwise disjoint cycles in S_n, then the order of the product $\alpha_1\alpha_2\ldots\alpha_m$ is equal to* $\operatorname{lcm}(|\alpha_1|, |\alpha_2|, \ldots, |\alpha_m|)$.

PROOF. Let $|\alpha_i| = k_i$ and let $k = \operatorname{lcm}(k_1, k_2, \ldots, k_m)$. By Lemma 2.6.1, the pairwise disjoint cycles commute. Therefore, $(\alpha_1\alpha_2\ldots\alpha_m)^k = \alpha_1^k\alpha_2^k\ldots\alpha_m^k = e$. Suppose $\ell > 0$ and $e = (\alpha_1\alpha_2\ldots\alpha_m)^\ell = \alpha_1^\ell\alpha_2^\ell\ldots\alpha_m^\ell$. The permutation $\alpha_2^\ell\ldots\alpha_m^\ell$ fixes point-wise every element of the orbit of α_1. Therefore, $\alpha_1^\ell = e$, hence $\ell \geq k_1$. By symmetry, $\ell \geq k_i$ for each i. □

COROLLARY 2.6.5. *Every $\pi \in S_n$ is a product of transpositions.*

PROOF. Let $k \geq 2$. By Theorem 2.6.3, it suffices to show that any k-cycle can be written as a product of transpositions. Notice that a 2-cycle (a_1a_2) is already a transposition, a 3-cycle $(a_1a_2a_3) = (a_1a_3)(a_1a_2)$ can be factored as a product of 2 transpositions, and a 4-cycle $(a_1a_2a_3a_4) = (a_1a_4)(a_1a_3)(a_1a_2)$ factors into 3 transpositions. In general, a k-cycle $(a_1a_2\cdots a_k) = (a_1a_k)\cdots(a_1a_3)(a_1a_2)$ can be written as a product of $k-1$ transpositions. □

6.2. The sign of a permutation. Let $n \geq 2$ and S_n the symmetric group on n letters. Let $x_1, \ldots, x_n$ be indeterminates and $\mathbb{Z}[x_1, \ldots, x_n]$ the ring of polynomials with coefficients in $\mathbb{Z}$. Given $\sigma \in S_n$, we define an automorphism $\sigma : \mathbb{Z}[x_1, \ldots, x_n] \to \mathbb{Z}[x_1, \ldots, x_n]$ by the rule $\sigma(p(x_1, \ldots, x_n)) = p(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. Since $\sigma(\tau(x_i)) = \sigma(x_{\tau(i)}) = x_{\sigma\tau(i)} = \sigma\tau(x_i)$, it follows that S_n acts as a group of permutations of $\mathbb{Z}[x_1, \ldots, x_n]$. Because $x_1, \ldots, x_n$ are indeterminates, it follows that σ defines an automorphism of the polynomial ring, hence we have a homomorphism of groups $S_n \to \operatorname{Aut}(\mathbb{Z}[x_1, \ldots, x_n])$. Now look at the polynomial

$$\Phi(x_1, \ldots, x_n) = \prod_{1 \leq i < j \leq n} (x_i - x_j).$$

Then Φ has degree $\binom{n}{2}$. Fix a transposition $\theta = (k, \ell)$ in S_n where $1 \leq k < \ell \leq n$. We compute $\theta(\Phi)$. If $\{i, j\} \cap \{k, \ell\} = \emptyset$, then $\theta(x_i - x_j) = x_i - x_j$. It is enough to consider terms with x_k or x_ℓ. All such terms except $(x_k - x_\ell)$ can be grouped into pairs. There are

four cases:

$$\begin{aligned}
&\theta((x_i-x_k)(x_i-x_\ell)) = (x_i-x_k)(x_i-x_\ell) && \text{if } i<k\\
&\theta((x_k-x_\ell)) = -(x_k-x_\ell) && \text{if } i=k \text{ (or } i=\ell)\\
&\theta((x_k-x_i)(x_i-x_\ell)) = (x_\ell-x_i)(x_i-x_k) = (x_k-x_i)(x_i-x_\ell) && \text{if } k<i<\ell\\
&\theta((x_k-x_i)(x_\ell-x_i)) = (x_\ell-x_i)(x_k-x_i) && \text{if } \ell<i
\end{aligned}$$

from which it follows that $\theta(\Phi) = -\Phi$. Therefore, if σ is written as a product of k transpositions, then $\sigma(\Phi) = (-1)^k\Phi$. The rule

$$\operatorname{sign}(\sigma) = \frac{\sigma(\Phi)}{\Phi}$$

defines a function $\operatorname{sign} : S_n \to \{1,-1\}$ which is an epimorphism of multiplicative groups. The kernel of the homomorphism $\operatorname{sign} : S_n \to \{1,-1\}$ is called the *alternating group on n letters* and is denoted denoted A_n. We return to the study of the alternating group in Section 2.6.4.

6.3. Conjugacy classes of the Symmetric Group. Let $n \geq 2$ and S_n the symmetric group on n letters. We view S_n as the group $\operatorname{Perm}(\mathbb{N}_n)$. The purpose of this section is to describe the conjugacy classes of S_n in terms of the partitions of the number n. If $\sigma \in S_n$, then we can write σ as a product of disjoint cycles $\sigma = \sigma_1\sigma_2\cdots\sigma_k$ where we assume $|\sigma_i| = s_i$ and $s_1 \geq s_2 \geq \cdots \geq s_k$. Furthermore, by adjoining 1-cycles if necessary, we assume $n = s_1 + s_2 + \cdots + s_k$. In other words, the sequence $s_1 \geq s_2 \geq \cdots \geq s_k$ is a partition of n. The next lemma shows that the conjugacy classes of S_n correspond to the partitions of n.

Let σ and θ be arbitrary permutations in S_n. Suppose $\sigma(i) = j$, $\theta(i) = k$, and $\theta(j) = \ell$. Then $\theta\sigma\theta^{-1}(k) = \theta\sigma(i) = \theta(j) = \ell$. This provides us with an algorithm to compute the cycle decomposition of the conjugation of σ by θ^{-1} given the cycle decomposition of σ: replace each letter by its image under θ. For instance, write $\sigma = \sigma_1\sigma_2\cdots\sigma_k$ as a product of disjoint cycles where $|\sigma_i| = s_i$, $s_1 \geq s_2 \geq \cdots \geq s_k$, and $n = s_1 + s_2 + \cdots + s_k$. Write $\sigma_i = (\sigma_{i1}, \sigma_{i2}, \ldots, \sigma_{is_i})$. Then $\theta\sigma_i\theta^{-1}$ is the cycle $(\theta(\sigma_{i1}), \theta(\sigma_{i2}), \ldots, \theta(\sigma_{is_i}))$. This shows that under conjugation the form of the cycle decomposition is preserved.

We illustrate this procedure by an example with $n = 10$. Let

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 3 & 8 & 4 & 5 & 1 & 10 & 9 & 7 & 6 & 2 \end{bmatrix}$$

$$\theta = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 5 & 4 & 10 & 1 & 7 & 3 & 9 & 8 & 6 & 2 \end{bmatrix}$$

Then

$$\theta\sigma\theta^{-1} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 7 & 4 & 2 & 8 & 10 & 3 & 5 & 9 & 6 & 1 \end{bmatrix}$$

As a product of disjoint cycles, we have $\sigma = (2,8,7,9,6,10)(1,3,4,5)$. Now compute the disjoint cycle form of the conjugate $\theta\sigma\theta^{-1}$. Because σ_1 starts with 2, and σ_2 starts with 1, we start the 6-cycle of $\theta\sigma\theta^{-1}$ with $\theta(2) = 4$, and the 4-cycle with $\theta(1) = 5$:

$$\begin{aligned}
\theta\sigma\theta^{-1} &= (4,8,9,6,3,2)(5,10,1,7)\\
&= \big(\theta(2),\theta(8),\theta(7),\theta(9),\theta(6),\theta(10)\big)\big(\theta(1),\theta(3),\theta(4),\theta(5)\big).
\end{aligned}$$

The last equation shows that the cycle decomposition can be obtained by applying θ to each letter in σ.

Now we show that every conjugacy class contains a canonical permutation. We continue to employ the notation established above. Consider the permutation

$$L = \begin{bmatrix} 1 & 2 & \dots & s_1 & s_1+1 & s_1+2 & \dots & s_1+s_2 & \dots & n \\ \sigma_{11} & \sigma_{12} & \dots & \sigma_{1s_1} & \sigma_{21} & \sigma_{22} & \dots & \sigma_{2s_2} & \dots & \sigma_{ks_k} \end{bmatrix}$$

where the second row is obtained by removing all of the parentheses from the product of disjoint cycles $\sigma_1\sigma_2\cdots\sigma_k$. Hence L is a permutation in S_n. Set $\tau = L^{-1}\sigma L$. Then the disjoint cycle decomposition of τ is obtained by inserting parentheses into $1,2,\dots,n$ and splitting it into cycles with the lengths $s_1,\dots,s_k$.

We illustrate this algorithm on the example from above. Start with the permutation $\sigma = (2,8,7,9,6,10)(1,3,4,5)$ in S_{10}. Then

$$L = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 8 & 7 & 9 & 6 & 10 & 1 & 3 & 4 & 5 \end{bmatrix}$$

is the permutation whose second row is obtained by removing the parentheses from σ. Compute:

$$L^{-1}\sigma L = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 2 & 3 & 4 & 5 & 6 & 1 & 8 & 9 & 10 & 7 \end{bmatrix}.$$

We see that $L^{-1}\sigma L = (1,2,3,4,5,6)(7,8,9,10)$ in disjoint cycle form.

The two algorithms specified above combine to prove Lemma 2.6.6.

LEMMA 2.6.6. *Let $n \geq 2$ and S_n the symmetric group on n letters. Two permutations σ,τ in S_n are in the same conjugacy class if and only if they give rise to the same partition of n. The number of distinct conjugacy classes of S_n is equal to the number of distinct partitions of n.*

6.4. The Alternating Group. Let $n \geq 3$. The alternating group on n letters is denoted A_n and is defined to be the kernel of the homomorphism $\text{sign} : S_n \to \{1,-1\}$. That is, A_n is the subgroup of all even permutations. We have $[S_n : A_n] = 2$ and $|A_n| = n!/2$. Theorem 2.6.9, the main result of this section, is a proof that if $n \neq 4$, then A_n is simple. The proof we give is completely elementary. In Exercise 2.6.12 the reader is asked to prove that A_4 contains a normal subgroup of order 4, hence A_4 is not simple.

LEMMA 2.6.7. *If $n \geq 3$, then A_n is generated by 3-cycles.*

PROOF. By Corollary 2.6.5, a 3-cycle is even, so A_n contains every 3-cycle. Every permutation in A_n is a product of an even number of transpositions. It suffices to show that a typical product $(ab)(cd)$ factors into 3-cycles. If (ab) and (cd) are disjoint, then we see that

$$\begin{aligned}(ab)(cd) &= (ab)(ac)(ac)(cd) \\ &= (acb)(acd)\end{aligned}$$

is a product of 3-cycles. If $a = c$, then we have $(ab)(ad) = (adb)$. These are the only cases, so A_n is generated by 3-cycles. □

LEMMA 2.6.8. *Let $n \geq 3$. If N is a normal subgroup of A_n and N contains a 3-cycle, then $N = A_n$.*

PROOF. Without loss of generality assume $(123) \in N$. Then $(123)(123) = (132) \in N$. We assume $n > 3$, otherwise we are done. By Corollary 2.6.5, a 3-cycle is even, so A_n contains every 3-cycle. Let $3 < a \leq n$ be arbitrary. We use the fact that $\sigma^{-1}N\sigma \subseteq N$ for all

$\sigma \in A_n$. Then $(1a3)(123)(13a) = (1a2)$ is in N. Also, $(1a2)^2 = (12a) \in N$. Similarly, we see that $(13a), (1a3), (23a), (2a3)$ are in N.

Now let $a \neq b$, $a > 2$, and $b > 2$. Then $(1b2)(12a)(12b) = (1ab)$ is in N. Similarly, we see that $(2ab), (3ab), (a1b), (a2b)$, etc. are in N.

Now let $a \neq b \neq c$, $a > 1$, $b > 1$, and $c > 1$. Then $(ac1)(a1b)(a1c) = (abc)$ is in N. So N contains every 3 cycle. By Lemma 2.6.7, $N = A_n$. □

THEOREM 2.6.9. *The alternating group A_n is simple if $n \neq 4$.*

PROOF. If $n = 2$, then $A_2 = \langle e \rangle$. If $n = 3$, then $A_3 = \langle (123) \rangle$ is a cyclic group of order 3, hence is simple. From now on assume $n > 4$, N is a normal subgroup of A_n and $N \neq \langle e \rangle$. We prove that $N = A_n$. The proof consists of a case-by-case analysis.

Case 1: If N contains a 3-cycle, then $N = A_n$, by Lemma 2.6.8.

Case 2: Assume N contains a permutation σ such that the cycle decomposition of σ has a cycle of length $r \geq 4$. Write $\sigma = (a_1 a_2 \cdots a_r)\tau$, where τ fixes each $a_1, \ldots, a_r$ element-wise. Let $\delta = (a_1 a_2 a_3)$. Then $\delta \in A_n$ and $\delta\sigma\delta^{-1} \in N$ since N is normal. The following computation

$$\begin{aligned}\sigma^{-1}\delta\sigma\delta^{-1} &= \tau^{-1}(a_1 a_r \cdots a_2)(a_1 a_2 a_3)(a_1 a_2 \cdots a_r)\tau(a_1 a_3 a_2) \\ &= (a_1 a_3 a_r)\end{aligned}$$

shows that Case 2 reduces to Case 1.

Case 3: Assume N has a permutation σ such that the cycle decomposition of σ has at least two disjoint 3-cycles. Write $\sigma = (a_1 a_2 a_3)(a_4 a_5 a_6)\tau$, where τ fixes each $a_1, a_2, a_3, a_4, a_5, a_6$ element-wise. Let $\delta = (a_1 a_2 a_4)$. Then $\delta \in A_n$ and $\delta^{-1}\sigma\delta \in N$ since N is normal. The following computation

$$\begin{aligned}\delta^{-1}\sigma\delta\sigma^{-1} &= (a_1 a_4 a_2)(a_1 a_2 a_3)(a_4 a_5 a_6)\tau(a_1 a_2 a_4)\tau^{-1}(a_1 a_3 a_2)(a_4 a_6 a_5) \\ &= (a_1 a_4 a_2 a_3 a_5)\end{aligned}$$

shows that Case 3 reduces to Case 2.

Case 4: Assume N has a permutation σ such that the cycle decomposition of σ consists of one 3-cycles and one or more 2-cycles. Write $\sigma = (a_1 a_2 a_3)\tau$, where τ is the product of the 2-cycles. Then $\sigma^2 = (a_1 a_3 a_2) \in N$, hence Case 4 reduces to Case 1.

Case 5: Assume every $\sigma \in N$ has a cycle decomposition that is a product of disjoint 2-cycles. Let $\sigma = (a_1 a_2)(a_3 a_4)\tau$ where τ is a product of 2-cycles and is disjoint from $(a_1 a_2)(a_3 a_4)$. Let $\delta = (a_1 a_2 a_3)$. Then $\delta \in A_n$ and $\delta^{-1}\sigma\delta \in N$ since N is normal. The following computation

$$\begin{aligned}\delta^{-1}\sigma\delta\sigma^{-1} &= (a_1 a_3 a_2)(a_1 a_2)(a_3 a_4)\tau(a_1 a_2 a_3)(a_1 a_2)(a_3 a_4)\tau \\ &= (a_1 a_4)(a_2 a_3)\end{aligned}$$

shows that $\beta = (a_1 a_4)(a_2 a_3)$ is in N. Since $n > 4$ (notice that this is the first time we have used this hypothesis), there exists $a_5 \notin \{a_1, a_2, a_3, a_4\}$. Let $\alpha = (a_1 a_4 a_5)$. The following computation

$$\begin{aligned}\alpha^{-1}\beta\alpha\beta &= (a_1 a_5 a_4)(a_1 a_4)(a_2 a_3)(a_1 a_4 a_5)(a_1 a_4)(a_2 a_3) \\ &= (a_1 a_4 a_5)\end{aligned}$$

shows that N contains a 3-cycle, hence Case 5 reduces to Case 1. □

COROLLARY 2.6.10. *If $n > 4$, the normal subgroups of S_n are $\langle e \rangle$, A_n, and S_n.*

PROOF. Let N be a normal subgroup of S_n. Then $N \cap A_n$ is a normal subgroup of A_n. By Theorem 2.6.9, $N \cap A_n$ is equal to either $\langle e \rangle$, or A_n. If $N \cap A_n = A_n$, then $[S_n : A_n] = 2$ implies $N = A_n$, or $N = S_n$. Suppose $N \cap A_n = \langle e \rangle$ and for contradiction's sake, suppose $N \neq \langle e \rangle$. Then N consists of e and odd permutations. If $\sigma \in N$ is an odd permutation, then σ^2 is even, hence $\sigma^2 \in N \cap A_n = \langle e \rangle$. Therefore, every element of N has order 2 or 1. Let $\sigma \in N$ and assume σ has order 2. By Corollary 2.6.4, σ decomposes into a product of disjoint transpositions. If $\sigma = (ab)$ is a transposition, then $(ab)(acb)(ab)(abc) = (acb)$ is in N, a contradiction. Assume $\sigma = (ab)(cd)\tau$, where τ is a product of disjoint transpositions that do not involve a, b, c, d. Let $\alpha = (acb)\sigma(abc) = (ac)(bd)\tau$. Then α is in N, and $\sigma\alpha = (ad)(bc)$ is in N. But $(ad)(bc)$ is even, which is a contradiction. □

COROLLARY 2.6.11. *Let $n > 4$. If H is a subgroup of S_n and $[S_n : H] < n$, then $H = A_n$ or $H = S_n$.*

PROOF. Let H be a subgroup of S_n, $m = [S_n : H]$, and assume $m < n$. Then S_n acts on G/H by left multiplication. If we identify $\mathrm{Perm}(G/H)$ with S_m, then there is a homomorphism of groups $\phi : S_n \to S_m$. By the Pigeonhole Principle (Exercise 1.1.11), $\ker\phi$ is a nontrivial normal subgroup of G. By Exercise 2.4.25, $\ker\phi$ is contained in H. By Corollary 2.6.10, $\ker\phi$ is either A_n or S_n. Therefore, H is either A_n or S_n. □

6.5. Exercises.

EXERCISE 2.6.12. Let $G = A_4$ be the alternating group on 4 letters. The order of G is twelve.

(1) Viewing G as a group of permutations of $\{1,2,3,4\}$, list the twelve elements of G using disjoint cycle notation. For each $x \in G$, compute the cyclic subgroup $\langle x \rangle$. Show that G has eight elements of order three and three elements of order two.
(2) Show that the subgroup of order 4 is the group of symmetries of a nonsquare rectangle (see Example 2.1.17).
(3) Show that G has four subgroups of order three. Show that the subgroup of order four is normal. Show that the center of G has order one. Construct the lattice of subgroups of G. Show that G has only one proper normal subgroup, namely the subgroup of order four.
(4) In Exercise 2.6.14 you are asked to compute the partition of G into conjugacy classes.

EXERCISE 2.6.13. As in Exercise 2.6.12, the alternating group on four letters is denoted A_4. Let N be the normal subgroup of A_4 of order four. Show that G is isomorphic to the semidirect product of N and a cyclic subgroup of order three that acts on N by conjugation.

EXERCISE 2.6.14. Let A_4 be the alternating group on 4 letters (see Exercise 2.6.12). Compute the partition of A_4 into conjugacy classes.

EXERCISE 2.6.15. Show that the set of transpositions $\{(12), (23), \ldots, (n-1, n)\}$ generates S_n.

EXERCISE 2.6.16. Show that S_n is generated by a transposition $(1,2)$ and an n-cycle $(123\cdots n)$.

EXERCISE 2.6.17. Compute the number of distinct k-cycles in S_n.

EXERCISE 2.6.18. Let $1 \leq k < n$. Show that for each k-subset $A = \{a_1, \ldots, a_k\}$ of $\mathbb{N}_n$ there is a subgroup of S_n isomorphic to $S_k \times S_{n-k}$. Show that any two such subgroups are conjugates of each other. (Hint: Suppose $a \in A$, $b \notin A$, and σ fixes $\mathbb{N}_n - A$. Look at $(ab)\sigma(ab)$.)

EXERCISE 2.6.19. Let $V = \{e, (12)(34), (13)(24), (14)(23)\}$ be the subgroup of order 4 in A_4. Show that V is a normal subgroup of S_4. Prove that S_4/V is a nonabelian group of order 6.

7. The Sylow Theorems

7.1. p-Groups. Let p be a prime number. A finite group G is called a *p-group* if $|G| = p^r$ for some $r \geq 1$. We begin this section with the following fundamental theorem on p-groups.

THEOREM 2.7.1. *(Fundamental Theorem on p-groups) Let p be a prime and G a finite group of order p^n, where $n \geq 1$. Then the following are true.*

(1) $Z(G) \neq \langle e \rangle$.
(2) If G has order p^2, then G is abelian.
(3) If $n > 1$, then G has a proper normal subgroup N such that $\langle e \rangle \neq N \neq G$.
(4) (A finite p-group is solvable) There is a sequence of subgroups $G_0 \subseteq G_1 \subseteq \cdots \subseteq G_{n-1} \subseteq G_n$ such that
  (a) $G_0 = \langle e \rangle$, $G_n = G$,
  (b) for $0 \leq i \leq n$, $|G_i| = p^i$,
  (c) for $0 \leq i \leq n-1$, G_i is a normal subgroup of G_{i+1} and the quotient G_{i+1}/G_i is a cyclic group of order p.

  We call $G_0, G_1, \ldots, G_n$ a solvable series *for G.*
(5) Let X be a finite set and assume G acts on X as a group of permutations. Let $X_0 = \{x \in X \mid g * x = x \text{ for all } g \in G\}$. Then $|X| \equiv |X_0| \pmod{p}$.

PROOF. (5): If $x \in X$, then $x \in X_0$ if and only if $G * x = \{x\}$. If $X_0 = X$, there is nothing to prove. Let $x_1, \ldots, x_m$ be a full set of representatives of the orbits with length two or more. The orbit decomposition of X is $X_0 \cup \left(\cup_{i=1}^m G * x_i\right)$. Taking cardinalities and applying Theorem 2.4.10,

$$\begin{aligned} |X| &= |X_0| + \sum_{i=1}^m |G * x_i| \\ &= |X_0| + \sum_{i=1}^m [G : G_{x_i}]. \end{aligned}$$

Then $[G : G_{x_i}] \neq 1$ for each i and by Corollary 2.2.12, $[G : G_{x_i}]$ divides p^n. Reducing both sides of the equation modulo p, we get $|X| \equiv |X_0| \pmod{p}$.

(1): Let G act on itself by conjugation. Then $Z(G)$ is the set of all elements fixed by the group action. By Part (5), $0 \equiv |Z(G)| \pmod{p}$.

(2): By Part (1), $Z(G)$ has order p or p^2. Then $G/Z(G)$ has order 1 or p, hence is cyclic. By Exercise 2.3.38, G is abelian.

(3): By Part (1), if $Z(G) \neq G$, then $N = Z(G)$ works. If $Z(G) = G$, then G is abelian. Every subgroup of G is abelian, so it suffices to find a proper subgroup of G. Let $z \in G - \langle e \rangle$ and set $N = \langle z \rangle$. If $G \neq N$, then we are done. Otherwise, $N = G$ and z has order p^n. By Lemma 2.2.16, $|z^p| = p^{n-1}$. In this case, $N = \langle z^p \rangle$ works.

(4): The proof is by induction on n. If $n = 1$, then $G_0 = \langle e \rangle$, $G_1 = G$ is a solvable series. If $n = 2$, then by Part (3) $G_0 = \langle e \rangle$, $G_1 = N$, $G_2 = G$ is a solvable series.

Inductively, assume $n \geq 2$ and that a solvable series exists for any p-group of order less than p^n. By Part (3) there exists a proper normal subgroup N. Then $|N| = p^t$, where $1 \leq t < n-1$. By our induction hypothesis, let $G_0 = \langle e \rangle, G_1, \ldots, G_t = N$ be a solvable series for N. Let $H = G/N$. By Corollary 2.2.12, $|H| = p^{n-t}$. By our induction hypothesis, let $H_0 = \langle e \rangle, H_1, \ldots, H_{n-t-1}, H_{n-t} = H$ be a solvable series for $H = G/N$. By Theorem 2.3.13, we lift each H_i to a subgroup G_{i+t} of G and get a sequence $G_t = N \subseteq G_{t+1} \subseteq \cdots \subseteq G_{n-1} \subseteq G_n = G$. By Theorem 2.3.12, $G_{i+1+t}/G_{i+t} \cong H_{i+1}/H_i$ for each $0 \leq i \leq t$. Combining the two sequences, $G_0 \subseteq \cdots \subseteq G_t \subseteq \cdots \subseteq G_{n-1} \subseteq G_n = G$ is a solvable series for G. □

LEMMA 2.7.2. *Let G be a finite group and p a prime number that divides $|G|$. If H is a subgroup of G and H is a p-group, then the following are true:*

(1) $[N_G(H) : H] \equiv [G : H] \pmod{p}$.
(2) If p divides $[G : H]$, then $[N_G(H) : H] > 1$ and $N_G(H) \neq H$.

PROOF. (1): As in Example 2.4.5, H acts on G/H by left multiplications: $h * xH = (hx)H$. Let $X = G/H$ and $X_0 = \{xH \in X \mid h * x = x \text{ for all } h \in H\}$. Then $xH \in X_0$ if and only if $x^{-1}hx \in H$ for all $h \in H$, which is true if and only if $x \in N_G(H)$. But $x \in N_G(H)$ if and only if $xH \subseteq N_G(H)$, hence X_0 consists of those cosets xH such that $xH \subseteq N_G(H)$. Then $|X_0| = [N_G(H) : H]$. By Theorem 2.7.1 (5), $|X| \equiv |X_0| \pmod{p}$, or $[G : H] \equiv [N_G(H) : H] \pmod{p}$.

(2): By Part (1), $0 \equiv [G : H] \equiv [N_G(H) : H] \pmod{p}$. Thus, $[N_G(H) : H]$ is a multiple of p. □

7.2. Cauchy's Theorem. The proof given below of Cauchy's Theorem is due to J. McKay [**10**]. This has been the proof of choice used in [**3**], [**8**], and other introductory texts on this subject.

THEOREM 2.7.3. *(Cauchy's Theorem) Let G be a finite group of order n and p a prime divisor of n. Then G contains an element of order p.*

PROOF. Let $X = G^p = \prod_{i=1}^{p} G$ be the product of p copies of G. Elements of G^p are p-tuples $(x_1, \ldots, x_p)$ where each x_i is in G and $|X| = n^p$. Let ξ be the p-cycle $(12\cdots p) \in S_p$. Then the cyclic subgroup $C = \langle \xi \rangle$ acts on X by

$$\xi^i * (x_1, \ldots, x_p) = \begin{cases} (x_p, x_1, \ldots, x_{p-1}) & \text{if } i = 1 \\ (x_{p-i+1}, \ldots, x_p, x_1, \ldots, x_{p-i}) & \text{if } 0 < i < p \\ (x_1, \ldots, x_p) & \text{if } i = 0 \text{ or } i = p. \end{cases}$$

Now define $Z = \{(x_1, \ldots, x_p) \in X \mid x_1 x_2 \cdots x_p = e\}$. Then Z is a subset of X. Given $x \in Z$, notice that $x_p = (x_1 \cdots x_{p-1})^{-1}$, so $|Z| = n^{p-1}$. Since $x_p = (x_1 \cdots x_{p-1})^{-1}$ implies $x_p x_1 x_2 \cdots x_{p-1} = e$, it follows that $\xi * Z = Z$. Hence C acts on Z and there is a partition of Z into orbits. Let Z_0 be the set of all z in Z fixed by ξ. A p-tuple $z = (x_1, \ldots, x_p)$ is fixed by ξ if and only if $x_1 = x_2 = \cdots = x_p$. Since $(e, e, \ldots, e) \in Z_0$, we know $Z_0 \neq \emptyset$. By Theorem 2.7.1 (5), $|Z_0| \equiv 0 \pmod{p}$. Then $|Z_0| \geq p$, and there are at least p elements $g \in G$ such that $g^p = e$. One solution to $g^p = e$ is $g = e$, any other solution is an element g of order p. □

7.3. The Sylow Theorems.

THEOREM 2.7.4. *(Sylow's First Theorem) Let G a finite group and p a prime number. If p^α divides $|G|$, then G contains a subgroup of order p^α.*

We give two proofs for Theorem 2.7.4. The first proof is due to H. Wielandt [**16**]. It has been the proof of choice used by [**3**], [**7**] and other introductory books on this subject.

FIRST PROOF OF THEOREM 2.7.4. Write $|G| = p^\gamma r$ where p^γ is the highest power of p that divides $|G|$. Then $0 \leq \alpha \leq \gamma$, and we write $|G| = p^\alpha q$. If we let $\beta = \gamma - \alpha$, then p^β is the highest power of p that divides q. Let X be the set of all subsets of G of cardinality p^α. Then

$$|X| = \binom{p^\alpha q}{p^\alpha} = \frac{p^\alpha q}{p^\alpha} \cdot \frac{p^\alpha q - 1}{p^\alpha - 1} \cdots \frac{p^\alpha q - i}{p^\alpha - i} \cdots \frac{p^\alpha q - p^\alpha + 2}{p^\alpha - p^\alpha + 2} \cdot \frac{p^\alpha q - p^\alpha + 1}{p^\alpha - p^\alpha + 1}$$

where the factorization on the right hand side results from expanding the binomial coefficient using Lemma 1.1.4, Let $0 < i < p^\alpha$ and write $i = p^t k$ where $0 \leq t < \alpha$ and $\gcd(p,k) = 1$. Then $p^\alpha q - i = p^t(p^{\alpha - t} q - k)$ and $p^{\alpha-t} q - k \equiv -k \pmod{p}$. This implies the highest power of p that divides $p^\alpha q - i$ is p^t. Therefore, canceling all powers of p from the numerator and denominator we see that the highest power of p that divides $|X|$ is the same as the highest power of p that divides q, which is p^β. As in Example 2.4.3, G acts on itself by left multiplication. If $a \in G$, and $S \in X$, then aS has cardinality p^α. Therefore, $a * S = aS$ defines an action by G on X. Under this action, X is partitioned into orbits. Since $p^{\beta+1}$ does not divide $|X|$, we know there is an orbit, say $G * S$, such that $p^{\beta+1}$ does not divide $|G * S|$, the length of the orbit. Let $H = G_S$ be the stabilizer of S. Then $H = \{h \in G \mid hS = S\}$. So $hs \in S$ for each $h \in H$ and $s \in S$. For a fixed $s \in S$, this implies the right coset Hs is a subset of S. Hence $|H| \leq |S| = p^\alpha$. By Corollary 2.2.12, $|G * S| = |G|/|H| = (p^\alpha q)/|H|$. Thus $p^\alpha q = |H||G * S|$. Since $p^{\alpha+\beta}$ divides the left hand side, we have $p^{\alpha+\beta}$ divides $|H||G * S|$. Since $p^{\beta+1}$ does not divide $|G * S|$, this implies p^α divides $|H|$. This proves H is a subgroup of G order p^α. □

SECOND PROOF OF THEOREM 2.7.4. Write $|G| = p^\gamma r$ where p^γ is the highest power of p that divides $|G|$. We prove more than is required. In fact, we show that G has a sequence of subgroups $P_0 \trianglelefteq P_1 \trianglelefteq \cdots \trianglelefteq P_\gamma$ such that $|P_i| = p^i$. Thus, this gives us a new proof of Theorem 2.7.1 (4). Set $P_0 = \langle e \rangle$, which has order 1. If $\gamma \geq 1$, then by Theorem 2.7.3, there exists $a \in G$ such that $P_1 = \langle a \rangle$ has order p. The method of proof is to iteratively apply Cauchy's Theorem $\gamma - 1$ times.

Inductively assume $1 \leq i < \gamma$, and that we have already constructed the sequence of subgroups $P_0 \trianglelefteq P_1 \trianglelefteq \cdots \trianglelefteq P_i$ in G, where $|P_i| = p^i$. To finish the proof it suffices to show that G has a subgroup P_{i+1} of order p^{i+1} containing P_i as a normal subgroup. By Corollary 2.2.12, $[G : P_i] = p^{\gamma - i} r$ is a multiple of p. By Lemma 2.7.2, $P_i \neq N_G(P_i)$ and p divides $[N_G(P_i) : P_i]$. Since P_i is normal in $N_G(P_i)$, by Theorem 2.7.3, the group $N_G(P_i)/P_i$. has a subgroup P'_{i+1} of order p. By Theorem 2.3.13, $P'_{i+1} = P_{i+1}/P_i$ for a subgroup P_{i+1} of $N_G(P_i)$ such that $P_i \subseteq P_{i+1} \subseteq N_G(P_i)$. By Corollary 2.2.12, $|P_{i+1}| = |P'_{i+1}||P_i| = p^{i+1}$. Since P_i is normal in $N_G(P_i)$, P_i is normal in P_{i+1}. □

By Theorem 2.7.4, if p is a prime, G is a finite group, $\alpha \geq 1$, and p^α is the highest power of p that divides $|G|$, then G has a subgroup of order p^α, call it P. In this case, we say P is a *p-Sylow subgroup* of G. Therefore, a p-Sylow subgroup is a maximal member of the set of all subgroups of G that are p-groups.

THEOREM 2.7.5. *(Sylow's Second Theorem) Let G be a finite group and p a prime that divides $|G|$. Then any two p-Sylow subgroups of G are conjugates of each other.*

PROOF. Assume G is not a p-group, otherwise there is nothing to prove. By Theorem 2.7.4, a p-Sylow subgroup exists. Let P and Q be two p-Sylow subgroups of G. We prove that there exists $x \in G$ such that $x^{-1}Px = Q$. Let $X = G/Q$ be the set of left cosets of Q in G. Let P act on X by left multiplication (Example 2.4.5). By Theorem 2.7.1 (5), $[G:Q] = |X| \equiv |X_0| \pmod{p}$. Since p does not divide $[G:Q]$, we know $|X_0| \neq 0$. Let $xQ \in X_0$. Then for each $a \in P$, $axQ = xQ$. Thus $x^{-1}ax \in Q$ for every $a \in P$, hence $x^{-1}Px \subseteq Q$. Since $|P| = |Q| = p^\alpha$, this implies $x^{-1}Px = Q$. □

COROLLARY 2.7.6. *Let G be a finite group and p a prime that divides $|G|$. Let P be a p-Sylow subgroup of G. Then the following are true.*

(1) For every $a \in G$, $a^{-1}Pa$ is a p-Sylow subgroup of G.
(2) In G, P is the unique p-Sylow subgroup if and only if P is a normal subgroup.
(3) $N_G(N_G(P)) = N_G(P)$.

PROOF. (1): Conjugation by a is an automorphism, hence $|P| = |a^{-1}Pa|$.

(2): The subgroup P is normal in G if and only if $P = a^{-1}Pa$ for all $a \in G$, which by (1) is true if and only if P is the unique p-Sylow subgroup of G.

(3): By Proposition 2.4.12, P is a normal subgroup of $N_G(P)$. By (2), P is the unique p-Sylow subgroup of $N_G(P)$. Let $z \in N_G(N_G(P))$. Then conjugation by z is an automorphism of $N_G(P)$, hence $zPz^{-1} = P$. This implies $z \in N_G(P)$. □

THEOREM 2.7.7. *(Sylow's Third Theorem) Let G be a finite group and p a prime that divides $|G|$. The number of p-Sylow subgroups in G is congruent to 1 modulo p and divides $|G|$. More precisely, let $|G| = p^\alpha r$ where $\alpha \geq 1$ and $\gcd(p,r) = 1$. If n is the number of p-Sylow subgroups in G, then n divides r and $n \equiv 1 \pmod{p}$.*

PROOF. By Theorem 2.7.4, a p-Sylow subgroup exists. Let P be a p-Sylow subgroup. As in Example 2.4.11, let G act by conjugation on 2^G, the power set of all subsets of G. By Theorem 2.7.5, the orbit of P is the set of all p-Sylow subgroups of G. The length of the orbit is $[G:N_G(P)]$, which divides $|G|$. But $r = [G:P] = [G:N_G(P)][N_G(P):P]$ shows the number of conjugates of P divides r.

Let X be the set of all p-Sylow subgroups of G. The number of p-Sylow subgroups in G is equal to $|X|$. Let P act on X by conjugation. By Theorem 2.7.1 (5), $|X| \equiv |X_0| \pmod{p}$. First note that $P \in X_0$. Suppose Q is another element of X_0. Then $a^{-1}Qa = Q$ for all $a \in P$. Therefore, $P \subseteq N_G(Q)$. In this case, both P and Q are p-Sylow subgroups of $N_G(Q)$. By Theorem 2.7.5, for some $x \in N_G(Q)$ we have $P = x^{-1}Qx$. But Q is normal in $N_G(Q)$, so $Q = x^{-1}Qx = P$. This proves $X_0 = \{P\}$. We have shown that $|X| \equiv 1 \pmod{p}$. □

PROPOSITION 2.7.8. *Let G be a finite group of order n where the unique factorization of n is $p_1^{e_1} \cdots p_m^{e_m}$. Assume for each p_i that G has a unique p_i-Sylow subgroup P_i. Then G is the internal direct product of $P_1, \ldots, P_m$.*

PROOF. By Corollary 2.7.6, each P_i is a normal subgroup of G. We use induction on m to show that $P_1, \ldots, P_m$ satisfy the criteria of Proposition 2.5.5 (4). If $m = 1$, there is nothing to prove. Assume $m > 1$. Then $P_{m-1}P_m$ is a subgroup of G because P_{m-1} is normal. Also, $P_{m-1} \cap P_m = \langle e \rangle$ by Lagrange's Theorem (Corollary 2.2.12), because $p_{m-1} \neq p_m$. Inductively assume $1 < r < m$ and that

(1) $P_{r+1}\cdots P_m$ is a subgroup of G, and
(2) for $i \in \{r,\ldots,m-1\}$, $P_i \cap (P_{i+1}\cdots P_m) = \langle e\rangle$.

Because P_{r-1} is normal in G, by Exercise 2.3.18, $P_{r-1}P_r\cdots P_m$ is a subgroup of G. The order of $P_r\cdots P_m$ is $p_r^{e_r}\cdots p_m^{e_m}$, by Lemma 2.5.4 (5). Because p_{r-1} is relatively prime to $|P_r\cdots P_m|$, by Lagrange's Theorem (Corollary 2.2.12), we know that $P_{r-1}\cap(P_r\cdots P_m) = \langle e\rangle$. By Mathematical Induction, this proves $P_1\cdots P_m$ is the internal direct product of $P_1,\ldots,P_m$. Since $|P_1\cdots P_m| = |G|$, this proves the proposition. □

EXAMPLE 2.7.9. Let p and q be distinct primes, and assume $p < q$. By Theorems 2.7.8 and 2.5.2, an abelian group of order pq is cyclic. If $q \equiv 1 \pmod p$, then by Theorem 2.3.27 there is a subgroup of order p in $\mathrm{Aut}(\mathbb{Z}/q) \cong U_q$. By Lemma 2.3.26, there exists a monomorphism $\theta : \mathbb{Z}/p \to \mathrm{Aut}(\mathbb{Z}/q)$. Using θ, the semidirect product $\mathbb{Z}/q \rtimes \mathbb{Z}/p$ is a nonabelian group of order pq. If q is not congruent to 1 modulo p, then by Theorem 2.7.7, we see that in a group of order pq every Sylow subgroup is normal, and a group of order pq is abelian.

We will prove in Corollary 3.6.10 that the group U_q is cyclic. Therefore, if $q \equiv 1 \pmod p$, then there is a unique subgroup of order p in $\mathrm{Aut}(Q)$. Therefore, the monomorphism θ is unique up to the choice of a generator for $\mathbb{Z}/p$. Hence there is at most one nonabelian group of order pq up to isomorphism.

7.4. Exercises.

EXERCISE 2.7.10. Let G be a finite group and N a normal subgroup of G. Show that if p is a prime and $|N| = p^r$ for some $r \geq 1$, then N is contained in every p-Sylow subgroup of G. See Exercise 2.7.13 for an application of this exercise.

EXERCISE 2.7.11. Let $n \geq 1$, A a nonempty set, and $X = A^n$ the product of n copies of A. An element x of X is an n-tuple $(x_1,\ldots,x_n)$ where each $x_i \in A$. Alternatively, an n-tuple $x = (x_1,\ldots,x_n)$ can be viewed as a function $x : \mathbb{N}_n \to A$ (see Section 1.1.3) where $x(i) = x_i$. Show that the symmetric group S_n acts on X by the rule $\sigma * x = x\sigma^{-1}$ where $x\sigma^{-1}$ refers to the composition of functions:

$$\mathbb{N}_n \xrightarrow{\sigma^{-1}} \mathbb{N}_n \xrightarrow{x} A.$$

EXERCISE 2.7.12. Let G be a group containing subgroups A and B such that $A \subseteq B \subseteq G$.

(1) Give an example such that B is normal in G, A is normal in B, and A is not normal in G. We say that normal over normal is not normal.
(2) Suppose G is finite and p is a prime number. Assume B is normal in G and A is normal in B and that A is a p-Sylow subgroup of B. Prove that A is normal in G.

EXERCISE 2.7.13. Let G be a group of order $2^r \cdot 7$, where $r \geq 5$. Apply Exercises 2.4.25 and 2.7.10 to show G contains a normal subgroup N satisfying: $2^{r-4} \leq |N| \leq 2^r$ and N is contained in every 2-Sylow subgroup of G.

EXERCISE 2.7.14. Let G be a finite group of order n.

(1) Show that for each n in the list: $30, 36, 40, 42, 44, 48, 50, 52, 54, 55, 56, 75, 3^2\cdot 5^2$, $9\cdot 37$, G is not a simple group.
(2) Show that for each n in the list: $45, 51, 5\cdot 17, 5^2\cdot 17, 5^2\cdot 37$, G is abelian.

EXERCISE 2.7.15. Let G be a group of order p^2q, where p and q are distinct primes. Show that G is not simple.

EXERCISE 2.7.16. Let G be a group of order $(p-1)p^2$, where p is an odd prime. Prove the following.

(1) G has a unique p-Sylow subgroup.
(2) There are at least four groups of order $(p-1)p^2$ which are pairwise non-isomorphic.

EXERCISE 2.7.17. Show that a group of order 105 is a semidirect product of two cyclic groups. Show how to construct an example of a nonabelian group of order 105.

8. Finite Abelian Groups

The purpose of this section is to prove that a finite abelian group can be decomposed into an internal direct product of cyclic subgroups in an essentially unique way. This is called the Basis Theorem for finite abelian groups.

8.1. The nth power map. Let A be an abelian group written multiplicatively and $n \in \mathbb{Z}$. The nth power map $\pi^n : A \to A$ is defined by the rule $\pi^n(x) = x^n$.

By Exercise 2.3.16 (where the abelian group was written additively) we see that π^n is an endomorphism of A with kernel $\{x \in A \mid |x| \text{ divides } n\}$ and image $\{x^n \mid x \in A\}$. In the following, the kernel of π^n will be denoted $A(n)$ and the image will be denoted A^n. Then $A(n)$ and A^n are subgroups of A. By the Isomorphism Theorem, Theorem 2.3.12 (a), ϕ induces an isomorphism $A/A(n) \cong A^n$.

LEMMA 2.8.1. *Let $\phi : A \to B$ be an isomorphism of abelian groups. Then for any $n \in \mathbb{Z}$, the following are true.*

(1) $\phi : A(n) \to B(n)$ is an isomorphism.
(2) $\phi : A^n \to B^n$ is an isomorphism.
(3) $\phi : A/A(n) \to B/B(n)$ is an isomorphism.
(4) $\phi : A/A^n \to B/B^n$ is an isomorphism.

PROOF. (1): Let $x \in A(n)$. Then $(\phi(x))^n = \phi(x^n) = \phi(e) = e$ implies $\phi(A(n)) \subseteq B(n)$. Given $y \in B(n)$, $y = \phi(x)$ for some $x \in A$. Then $e = y^n = (\phi(x))^n = \phi(x^n)$. So $x \in \ker(\phi) = \langle e \rangle$. This proves $\phi : A(n) \to B(n)$ is an isomorphism.

(2): Let $x \in A$. Then $\phi(x^n) = (\phi(x))^n$, so $\phi(A^n) \subseteq B^n$. Let $y^n \in B^n$. Then $y = \phi(x)$ for some $x \in A$, so $y^n = (\phi(x))^n = \phi(x^n)$, which proves $\phi : A^n \to B^n$ is an isomorphism.

(3): Consider the commutative diagram

$$\begin{array}{ccc} A & \xrightarrow{\phi} & B \\ \downarrow & & \downarrow \eta \\ A/\ker(\eta\phi) & \xrightarrow{\cong} & B/B(n) \end{array}$$

where all of the maps are onto. By Part (1), the kernel of $\eta\phi$ is $\phi^{-1}(B(n)) = A(n)$. By Theorem 2.3.12 (a), $\eta\phi$ factors through $A/A(n)$ giving the isomorphism: $A/A(n) \cong B/B(n)$.

(4): Consider the commutative diagram

$$\begin{array}{ccc} A & \xrightarrow{\phi} & B \\ \downarrow & & \downarrow \eta \\ A/\ker(\eta\phi) & \xrightarrow{\cong} & B/B^n \end{array}$$

where all of the maps are onto. By Part (2), the kernel of $\eta\phi$ is $\phi^{-1}(B^n) = A^n$. By Theorem 2.3.12 (a), $\eta\phi$ factors through A/A^n giving the isomorphism: $A/A^n \cong B/B^n$. $\square$

LEMMA 2.8.2. *Let $A = \langle a \rangle$ be an infinite cyclic group and $n \in \mathbb{N}$. Then $A(n) = \langle e \rangle$ and A/A^n is cyclic of order n.*

PROOF. We have the isomorphism $\phi : \mathbb{Z} \to A$ which is defined on generators by the rule $\phi(1) = a$ (Theorem 2.3.25 (5)). The group $\mathbb{Z}$ is written additively as in Exercise 2.3.16, and instead of the nth power map π^n, we will use the "left multiplication by n" map $\lambda_n : \mathbb{Z} \to \mathbb{Z}$. The kernel of λ_n is $\langle 0 \rangle$ and the image of λ_n is $\langle n \rangle = n\mathbb{Z}$. Applying Lemma 2.8.1 we have $A(n) = \langle e \rangle$ and $A/A^n \cong \mathbb{Z}/n\mathbb{Z}$ is cyclic of order n. □

LEMMA 2.8.3. *Let $A = \langle a \rangle$ be a finite cyclic group of order m and $n \in \mathbb{N}$. If $d = \gcd(m,n)$, then the following are true.*

(1) $A(n) = \langle a^{m/d} \rangle$ is cyclic of order d.
(2) $A/A(n) \cong A^n$ is cyclic of order m/d.
(3) A/A^n is cyclic of order d.

PROOF. We have $A = \{e, a, \ldots, a^{m-1}\}$.

(1): Suppose $0 \leq i < m$ and $(a^i)^n = e$. Then m divides ni and by Proposition 1.2.10, $\text{lcm}(m,n) = mn/d$ divides ni. This implies m/d divides i. Hence $A(n) \subseteq \langle a^{m/d} \rangle$. But $a^{m/d}$ has order d by Lemma 2.2.16. Since d divides n, $A(n) \supseteq \langle a^{m/d} \rangle$, proving (1).

(2) and (3): By Theorem 2.3.12 (a), $A/A(n) \cong A^n$. From Part (1) and Lagrange's Theorem (Corollary 2.2.12), we get (2). From Part (2) and Lagrange's Theorem, we get (3). □

LEMMA 2.8.4. *Let A and B be abelian groups and $n \in \mathbb{Z}$. Then the following are true.*

(1) $(A \times B)(n) = A(n) \times B(n)$.
(2) $(A \times B)^n = A^n \times B^n$.

PROOF. Let (a,b) be a typical element in $A \times B$. Part (2) follows immediately from the identity $(a,b)^n = (a^n, b^n)$. Part (1) follows from $(A \times B)(n) = \{(a,b) \mid (a,b)^n = (e,e)\} = \{(a,b) \mid a^n = e \text{ and } b^n = e\} = A(n) \times B(n)$. □

LEMMA 2.8.5. *Let A be a finite abelian group, p a prime, $r \in \mathbb{N}$, and assume p^r is the highest power of p that divides $|A|$. Then $A(p^r)$ is equal to the p-Sylow subgroup of A.*

PROOF. Since A is abelian, every subgroup is normal and by Corollary 2.7.6, A has a unique p-Sylow subgroup. Call it P. Then $|P| = p^r$. If $x \in P$, then $|x|$ divides p^r by Corollary 2.2.17. As a set, $A(p^r)$ consists of those elements $x \in A$ whose order divides p^r. Therefore, $P \subseteq A(p^r)$. If $x \in A(p^r)$, then by Exercise 2.7.10, x is in P. Therefore, $A(p^r) \subseteq P$. □

8.2. The Basis Theorem.

THEOREM 2.8.6. *Every finite abelian group G is isomorphic to an internal direct product of cyclic groups.*

PROOF. Since G is abelian, every subgroup of G is normal. It follows from Proposition 2.7.8 that G is isomorphic to the internal direct product of its Sylow subgroups. Therefore, it suffices to prove the theorem for a finite p-group. From now on, assume p is a prime and $[G : e] = p^n$, for some $n \in \mathbb{N}$.

The proof is by Mathematical Induction on n. If $n = 1$, then $G \cong \mathbb{Z}/p$ is cyclic. Assume inductively that $n > 1$ and that the theorem is true for all abelian groups of order p^i where $0 < i < n$.

Let $a \in G$ be an element of maximal order. If $|a| = p^n$, then $G = \langle a \rangle$ is cyclic and we are done. Assume $|a| = p^\alpha$, where $1 \leq \alpha < n$. Set $A = \langle a \rangle$. Look at the quotient G/A. We have $|G/A| = [G : A] = p^{n-\alpha}$. By our induction hypothesis, G/A is an internal direct product of cyclic groups. That is, there exist $b_1, \ldots, b_m \in G$ such that

$$G/A = \langle [b_1] \rangle \times \cdots \times \langle [b_m] \rangle \tag{8.1}$$

where we write $[b_i]$ for the left coset b_iA. Assume the order of $[b_i]$ in G/A is p^{β_i}. By Exercise 2.3.40, p^{β_i} divides the order of b_i in G . Since $|a|$ is maximal, $\alpha \geq \beta_i$ for each i. Because $(b_iA)^{p^{\beta_i}} = A$, $b_i^{p^{\beta_i}} \in A$. Therefore $b_i^{p^{\beta_i}} = a^{k_i}$ for some k_i. Because the order of every element of G divides p^α, we have

$$\left(a^{k_i}\right)^{p^{\alpha-\beta_i}} = \left(b_i^{p^{\beta_i}}\right)^{p^{\alpha-\beta_i}} = b_i^{p^\alpha} = e.$$

It follows that p^α divides $k_i p^{\alpha-\beta_i}$. Hence p^{β_i} divides k_i. Write $k_i = \ell_i p^{\beta_i}$. Set $a_i = b_i a^{-\ell_i}$. Then

$$a_i^{p^{\beta_i}} = \left(b_i a^{-\ell_i}\right)^{p^{\beta_i}} = b_i^{p^{\beta_i}} a^{-\ell_i p^{\beta_i}} = a^{k_i} a^{-k_i} = e$$

which implies $|a_i| \leq p^{\beta_i}$. Set $A_i = \langle a_i \rangle$. To finish the proof, we show that G is the internal direct product of $A, A_1, \ldots, A_m$. Let $x \in G$ be an arbitrary element of G. In G/A we can write the coset xA as a product $b_1^{e_1} A \cdots b_m^{e_m} A$. Since $b_iA = a_iA$, we see that $x = a_1^{e_1} \cdots a_m^{e_m} a^{e_0}$, for some $e_0 \in \mathbb{Z}$. This proves that $G = AA_1 \cdots A_m$.

Suppose $e = a^{e_0} a_1^{e_1} \cdots a_m^{e_m}$. In G/A we have $[e] = [a_1]^{e_1} \cdots [a_m]^{e_m} = [b_1]^{e_1} \cdots [b_m]^{e_m}$. As in Eq. (8.1), G/A is a direct product so $[b_i]^{e_i} = [e]$ for each i. So p^{β_i} divides e_i for each i. Therefore, $a_i^{e_i} = e$ for each i. It follows that $e = a^{e_0}$, hence e has a unique representation. □

THEOREM 2.8.7. *(Basis Theorem for Finite Abelian Groups) Let G be an abelian group of finite order. Then the following are true.*

1. *G is the internal direct product of its Sylow subgroups.*
2. *If p is a prime factor of $|G|$ and P is the unique p-Sylow subgroup of G, then there exist $a_1, \ldots, a_m$ in P such that P is the internal direct product of the cyclic subgroups $\langle a_1 \rangle, \ldots, \langle a_m \rangle$, the order of a_i is equal to p^{e_i}, and $e_1 \geq e_2 \geq \cdots \geq e_m$.*
3. *G is uniquely determined by the prime factors p of $|G|$ and the integers e_i that occur in (2).*

The prime powers p^{e_i} that occur in (3) are called the invariants *of G. Notice that if $|P| = p^n$, then $n = e_1 + \cdots + e_m$ is a partition of the integer n.*

PROOF. Part (1) follows from Proposition 2.7.8. Part (2) follows from Theorem 2.8.6.

(3): Let A and B be finite abelian groups. First we prove that if $\phi : A \to B$ is an isomorphism, then A and B have the same invariants. Because ϕ is a one-to-one correspondence, $|A| = |B|$. Let p be a prime that divides $|A|$ (and $|B|$). By Lemmas 2.8.5 and 2.8.1, the p-Sylow subgroups of A and B are isomorphic. Using Theorem 2.8.6 we can suppose the p-Sylow subgroup of A is the internal direct product of $A_1, \ldots, A_m$ where $A_i = \langle a_i \rangle$, $|a_i| = p^{e_i}$, and $e_1 \geq e_2 \geq \cdots \geq e_m \geq 1$. Likewise, assume the p-Sylow subgroup of B is the internal direct product of $B_1, \ldots, B_n$ where $B_i = \langle b_i \rangle$, $|b_i| = p^{f_i}$, and $f_1 \geq f_2 \geq \cdots \geq f_n \geq 1$. We have $A_1 \times \cdots \times A_m \cong B_1 \times \cdots \times B_n$. Multiply by p and apply Lemmas 2.8.1, 2.8.3 and 2.8.4 to get $(A_1 \times \cdots \times A_m)(p) \cong A_1(p) \times \cdots \times A_m(p)$ is a direct product of cyclic groups of order p, has order p^m, and is isomorphic to $(B_1 \times \cdots \times B_n)(p) \cong B_1(p) \times \cdots \times B_n(p)$ which has order p^n. Therefore $m = n$. Inductively, assume the uniqueness claim is true for any finite

p-group of order less than $p^{e_1+\cdots+e_m}$. By Lemma 2.8.3, the invariants of $(A_1\times\cdots\times A_m)^p = A_1^p\times\cdots\times A_m^p$ are $e_1-1\geq\cdots\geq e_m-1$ and the invariants of $(B_1\times\cdots\times B_m)^p = B_1^p\times\cdots\times B_m^p$ are $f_1-1\geq\cdots\geq f_m-1$. By induction, $e_i=f_i$ for each i.

For the converse, suppose we are given the cyclic groups $A_1,\ldots,A_m$, $B_1,\ldots,B_n$, where $|A_i|=p^{e_i}$ for each i, and $|B_j|=p^{f_i}$ for each j. If $m=n$ and $e_i=f_i$ for each i, then clearly $A_i\cong B_i$ for each i and we have $A_1\times\cdots\times A_m\cong B_1\times\cdots\times B_m$. □

8.3. Exercises.

EXERCISE 2.8.8. If G is any group, and $n\in\mathbb{N}$, the direct product of n copies of G is $G^n=\prod_{i=1}^n G$. Let $G,+$ be an abelian group. Using Exercise 2.3.16, show that an n-tuple $A\in(a_1,\ldots,a_n)\in\mathbb{Z}^n$ defines a homomorphism $A:G^n\to G$ by the rule $A(x_1,\ldots,x_n)=\sum_{i=1}^n a_ix_i$.

EXERCISE 2.8.9. Let $m,n\in\mathbb{N}$. Show that the direct product $\mathbb{Z}/m\times\mathbb{Z}/n$ is cyclic if and only if $\gcd(m,n)=1$.

EXERCISE 2.8.10. Let G be a finite abelian group. Prove that the following are equivalent:

(1) G is cyclic.
(2) For every prime factor p of $|G|$, the p-Sylow subgroup of G is cyclic.
(3) For every prime factor p of $|G|$, $G(p)$ (see Exercise 2.3.16 for this notation) is cyclic.
(4) For every $n\in\mathbb{N}$, the order of $G(n)$ is at most n.
(5) For every $n\in\mathbb{N}$, the equation $x^n=e$ has at most n solutions in G.

EXERCISE 2.8.11. Let A and B be abelian groups written additively. The set of all homomorphisms from A to B is denoted $\mathrm{Hom}(A,B)$.

(1) If $f,g\in\mathrm{Hom}(A,B)$, then $f+g$ is the function defined by the rule: $(f+g)(x)=f(x)+g(x)$. Show that this additive binary operation makes $\mathrm{Hom}(A,B)$ into an abelian group.
(2) Now consider the case where $A=B$. Show that composition of functions defines a binary operation on $\mathrm{Hom}(A,A)$ satisfying the following.
   (a) $f(gh)=(fg)h$ for all f,g,h in $\mathrm{Hom}(A,A)$. In other words, composition of functions is associative.
   (b) $f(g+h)=fg+fh$ and $(f+g)h=fh+gh$ for all f,g,h in $\mathrm{Hom}(A,A)$. In other words, composition distributes over addition.

   Together with the two binary operations of addition and composition of endomorphisms, we call $\mathrm{Hom}(A,A)$ the *ring of endomorphism of* A.

EXERCISE 2.8.12. Let $m,n\in\mathbb{N}$ be positive integers. Show that the abelian group $\mathrm{Hom}(\mathbb{Z}/m,\mathbb{Z}/n)$ is a cyclic group of order $\gcd(m,n)$. (Hints: Exercises 2.8.11 and 2.4.20.)

EXERCISE 2.8.13. If p is a prime, and $n\geq 1$, compute the following:

(1) Let $G=\prod_{i=1}^n\mathbb{Z}/2=\mathbb{Z}/2\times\cdots\times\mathbb{Z}/2$ be the direct product of n copies of $\mathbb{Z}/2$. How many subgroups of order 2 are there in G?
(2) Let $G=\prod_{i=1}^n\mathbb{Z}/p=\mathbb{Z}/p\times\cdots\times\mathbb{Z}/p$ be the direct product of n copies of $\mathbb{Z}/p$. How many elements of order p are there in G? How many subgroups of order p are there in G?
(3) Let $G=\prod_{i=1}^n\mathbb{Z}/p^{e_i}=\mathbb{Z}/p^{e_1}\times\cdots\times\mathbb{Z}/p^{e_n}$ where $e_i\geq 1$ for each i. How many elements of order p are there in G? How many subgroups of order p are there in G?

EXERCISE 2.8.14. Show that if G is a finite group of order at least three, then $\mathrm{Aut}(G)$ has order at least two.

9. Classification of Finite Groups

This section consists of computations and applications of the theorems from the previous sections. The examples presented here are not only intended to classify all groups of a given order, but to illustrate the various theorems of Group Theory.

9.1. Groups of order 12. We show in this example that up to isomorphism there are exactly five groups of order 12. Let G be a finite group of order $12 = 2^2 \cdot 3$. Let P be a 2-Sylow subgroup. Then P is either $\langle a \mid a^4 = e\rangle$, a cyclic group of order 4, or P is $\langle a,b \mid a^2 = b^2 = e, ab = ba\rangle$, an isomorphic copy of the Klein four group. In both cases P is abelian. By Theorem 2.7.7, the number of conjugates of P is odd and divides 3, hence P has either 1 or 3 conjugates. Let Q be a 3-Sylow subgroup. By Theorem 2.7.7, the number of conjugates of Q divides 4, hence Q has either 1 or 4 conjugates. We know that $Q = \langle c \mid c^3 = e\rangle$ is cyclic, hence abelian. Since $P \cap Q = \langle e\rangle$, by Theorem 2.2.14 we see that $PQ = G$. We consider the following four cases.

Case 1: Assume P and Q are both normal in G. By Theorem 2.7.8, G is the internal direct product of P and Q, hence G is abelian. By Theorem 2.8.7, G is isomorphic to either

$$\mathbb{Z}/3 \times \mathbb{Z}/4$$

or

$$\mathbb{Z}/3 \times \mathbb{Z}/2 \times \mathbb{Z}/2.$$

Case 2: Assume P is normal and Q has 4 conjugates. Then Q acts by conjugation on P and there is a homomorphism $\theta : Q \to \mathrm{Aut}(P)$, where $\theta(c) = \alpha_{c^{-1}}$ is conjugation by c^{-1}. By Exercise 2.4.18, G is isomorphic to $P \rtimes Q$, the semidirect product of P and Q.

There are two subcases to consider. If $P = \langle a\rangle$ is cyclic, then $\mathrm{Aut}(P) \cong U_4$ is a group of order two. Since Q has order three, in this case $\mathrm{im}\,\theta = \langle e\rangle$. Then $cac^{-1} = a$, hence G must be abelian. In this case, G is the first group of Case 1. If P is $\langle a,b \mid a^2 = b^2 = e, ab = ba\rangle$, then $\mathrm{Aut}(P)$ is isomorphic to $\mathrm{GL}_2(\mathbb{Z}/2)$. We will prove this in Proposition 4.5.8. By Example 2.1.20, $\mathrm{GL}_2(\mathbb{Z}/2) \cong S_3$. There are two elements of order 3 in S_3. One element of order three in $\mathrm{Aut}(P)$ is the cyclic permutation π defined by $a \mapsto b \mapsto ab \mapsto a$. The other element of order three is π^{-1}. Therefore, if $\theta(c) = \pi$, then $\theta(c^{-1}) = \pi^{-1}$. Since Q is generated by either c, or c^{-1}, without loss of generality we assume $\theta(c) = \pi$. Then $cac^{-1} = b$ and $cbc^{-1} = ab$. The semidirect product $P \rtimes Q$ has presentation in terms of generators and relations

$$\langle a,b,c \mid a^2 = b^2 = c^3 = e, ab = ba, cac^{-1} = b, cbc^{-1} = ab\rangle.$$

This group is isomorphic to A_4 by the map defined by $a \mapsto (12)(34)$, $b \mapsto (14)(23)$, $c \mapsto (123)$. The reader should verify that $(123)(12)(34)(132) = (14)(23)$, $(123)(14)(23)(132) = (13)(24)$, and $(123)(13)(24)(132) = (12)(34)$.

Case 3: Assume P has 3 conjugates and Q is normal. Then P acts on Q by conjugation and there is a homomorphism $\theta : P \to \mathrm{Aut}(Q)$. Then G is the semidirect product $Q \rtimes P$. By Theorem 2.3.27, $\mathrm{Aut}(Q) \cong U_3$ is a group of order 2. The automorphism of order two is defined by $c \mapsto c^{-1}$. There are two subcases to consider. If $P = \langle a\rangle$ is cyclic, then there is one nontrivial possibility for θ. In this case, $aca^{-1} = c^{-1}$. The presentation of the semidirect product in terms of generators and relations is

$$\langle a,c \mid a^4 = c^3 = e, aca^{-1} = c^{-1}\rangle.$$

If P is $\langle a,b \mid a^2 = b^2 = e, ab = ba\rangle$, then there are three subgroups of order two, hence three possible homomorphisms from P onto $\mathrm{Aut}(Q)$. Therefore, one of a, b, ab commutes with c. Since P is generated by any two of the three, without loss of generality we assume $aca = c^{-1}$ and $bcb = c$. The semidirect product is described by

$$\langle a,b,c \mid a^2 = b^2 = c^3 = e, ab = ba, aca = c^{-1}, bc = cb\rangle.$$

This group is isomorphic to D_6 the element bc has order 6, and $a(bc)a = (bc)^{-1}$. Another way to view this group is as the internal direct product $\langle b\rangle \times \langle a,c\rangle$ which is isomorphic to $\mathbb{Z}/2 \times D_3$.

Case 4: Assume P has 3 conjugates and Q has 4 conjugates. Counting elements we find that each subgroup of order 3 has 2 elements of order 3. Therefore, G has 8 elements of order 3. The subgroup P has 4 elements. Since P is not normal, the group G has more than 12 elements, which is a contradiction. Case 4 cannot occur.

9.2. Groups of order 30. In this example we show that up to isomorphism there are exactly 4 groups of order 30. Let G be a group of order $30 = 2 \cdot 3 \cdot 5$. Using Theorems 2.7.8 and 2.5.2 we see that if G is abelian, then G is cyclic. Let P be a 2-Sylow subgroup of G, Q a 3-Sylow subgroup, and R a 5-Sylow subgroup. By Theorem 2.7.7, Q is either normal or has 10 conjugates. The number of conjugates of R is either 1 or 6. By counting elements, we see that if G has 6 subgroups of order 5 then there are 24 elements of order 5. If G has 10 subgroups of order 3, then this includes 20 elements of order 3. Since $|G| = 30$, this implies either Q is normal or R is normal. By Exercise 2.3.18, QR is a subgroup of G. Since $Q \cap R = \langle e\rangle$, by Theorem 2.2.14, $|QR| = 15$. Since $[G : QR] = 2$, Exercise 2.3.17, implies QR is normal in G. By Theorem 2.5.2, QR is cyclic. Write $QR = \langle b\rangle$. Then P acts by conjugation on QR and there is a homomorphism $\theta : P \to \mathrm{Aut}(QR) \cong U_{15}$. The image of θ has order 1 or 2. The group U_{15} has order $\phi(15) = 8$. The reader should verify that there are 4 elements in U_{15} that satisfy $x^2 \equiv 1 \pmod{15}$, they are $1, 4, -1, -4$. Therefore, if $P = \langle a\rangle$, then $aba = b^s$, where $s \in \{1,4,-1,-4\}$. Thus G is the semidirect product $QR \rtimes P$. The presentation in terms of generators and relations is

$$G = \langle a,b \mid a^2 = b^{15} = e, aba = b^s\rangle \tag{9.1}$$

where $s \in \{1,4,-1,-4\}$. If $s = 1$, then a commutes with b, and G is abelian. If $s = -1$, then G is isomorphic to D_{15}. By Example 2.3.32, the center of D_{15} is $\langle e\rangle$.

If $s = 4$, then because $ab^5a = b^{20} = b^5$ we see that the center of G contains b^5, an element of order 3. Then $G/\langle b^5\rangle$ has presentation $\langle a,b \mid a^2 = b^5 = e, aba = b^4\rangle$ which is isomorphic to D_5. Since the center of D_5 is trivial, this proves the center of G is $Z = \langle b^5\rangle$. Since $ab^3a = b^{12} = b^{-3}$ we see that the subgroup $D = \langle a, b^3\rangle$ has order 10 and is isomorphic to D_5. generated by a and b^3. Using Exercise 2.5.19, we see that G is the internal direct product $D \times Z$, hence G is isomorphic to $D_5 \times \mathbb{Z}/3$.

If $s = -4$, then because $ab^3a = b^{-12} = b^3$ we see that the center of G contains b^3, an element of order 5. Then $G/\langle b^3\rangle$ has presentation $\langle a,b \mid a^2 = b^3 = e, aba = b^{-1}\rangle$ which is isomorphic to D_3. Since the center of D_3 is trivial, this proves the center of G is $Z = \langle b^3\rangle$. Since $ab^5a = b^{-20} = b^{-5}$ we see that the subgroup $D = \langle a, b^5\rangle$ has order 6 and is isomorphic to D_3. Using Exercise 2.5.19, we see that G is the internal direct product $D \times Z$, hence G is isomorphic to $D_3 \times \mathbb{Z}/5$.

This proves that in (9.1) the four values of s give rise to four groups that are pairwise nonisomorphic.

9.3. Groups of order 63. We show in this example that up to isomorphism there are exactly four groups of order 63. Let G be a finite group of order $63 = 7 \cdot 3^2$. If G is abelian, then by Theorem 2.8.7, G is isomorphic to either $\mathbb{Z}/7 \times \mathbb{Z}/9$, or $\mathbb{Z}/7 \times \mathbb{Z}/3 \times \mathbb{Z}/3$. Assume from now on that G is nonabelian. Let P be a 7-Sylow subgroup. The number of conjugates of P divides 9 and is of the form $1+7k$. Therefore, we conclude that $k = 0$ and P is normal. Let Q be a 3-Sylow subgroup. We know that Q is abelian. Since $P \cap Q = \langle e \rangle$, by Theorem 2.2.14 we see that $PQ = G$. By Exercise 2.4.18, $G = P \rtimes Q$ and the action by Q on P is conjugation. By Example 2.4.8, the homomorphism

$$\theta : Q \to \mathrm{Aut}(P) \cong U_7$$

is defined by $\theta(x) = \alpha_{x^{-1}}$, where $\alpha_{x^{-1}}$ is the inner automorphism of P corresponding to conjugation by x^{-1}. If the image of θ is $\langle 1 \rangle$, then every element of Q commutes with every element of P and G is abelian. By our assumption, we can assume θ is not the trivial map. By Theorem 2.3.27, $\mathrm{Aut}(P) \cong U_7$ which is an abelian group of order $\phi(7) = 6$, hence is cyclic. Since Q has order 9, this implies $\ker(\theta)$ has order 3, and $\mathrm{im}(\theta)$ has order 3. Let $P = \langle a \rangle$. There are two cases.

Case 1: $Q = \langle b \rangle$ is cyclic. Then θ maps b to $\alpha_{b^{-1}}$, the inner automorphism defined by b^{-1}, which is an element of order 3 in U_7. There are two elements of order 3 in U_7, namely [2] and [4]. Therefore, $bab^{-1} = a^i$ where $i = 2$ or 4. Notice that $|b^2| = 9$ so $Q = \langle b^2 \rangle$. Since $b^2ab^{-2} = a^{2i}$, without loss of generality we can replace b with b^2 if necessary and assume $i = 2$. Then in this case,

$$G = \langle a, b \mid a^7 = b^9 = e, bab^{-1} = a^2 \rangle$$

is the presentation of G in terms of generators and relations.

Case 2: Q is a direct sum of two cyclic groups of order 3. Suppose $\ker(\theta) = \langle c \rangle$ and $b \in Q - \langle c \rangle$. Then $Q = \langle b, c \rangle$. As in Case 1, $bab^{-1} = a^i$ where $i = 2$ or 4. Again, we can replace b with b^{-1} if necessary and assume $bab^{-1} = a^2$. Then in this case,

$$G = \langle a, b, c \mid a^7 = b^3 = c^3 = e, bc = cb, bab^{-1} = a^2, cac^{-1} = a \rangle$$

is the presentation of G.

For a continuation of this example, see Exercise 2.9.1.

9.4. Groups of order 171. We show in this example that up to isomorphism there are exactly five groups of order 171. Let G be a finite group of order $171 = 19 \cdot 3^2$. If G is abelian, then by Theorem 2.8.7, G is isomorphic to either $\mathbb{Z}/19 \times \mathbb{Z}/9$, or $\mathbb{Z}/19 \times \mathbb{Z}/3 \times \mathbb{Z}/3$. Assume from now on that G is nonabelian. Let P be a 19-Sylow subgroup. Then $P = \langle a \rangle$ is cyclic. The number of conjugates of P divides 9 and is of the form $1+19k$. Therefore, we conclude that $k = 0$ and P is normal. Let Q be a 3-Sylow subgroup. We know that Q is abelian. Since $P \cap Q = \langle e \rangle$, by Theorem 2.2.14 we see that $PQ = G$. By Exercise 2.4.18, $G = P \rtimes Q$ and the action by Q on P is conjugation. By Example 2.4.8, the homomorphism

$$\theta : Q \to \mathrm{Aut}(P) \cong U_{19}$$

is defined by $\theta(x) = \alpha_{x^{-1}}$, where $\alpha_{x^{-1}}$ is the inner automorphism of P corresponding to conjugation by x^{-1}. If the image of θ is $\langle 1 \rangle$, then every element of Q commutes with every element of P and G is abelian. By our assumption, we can assume θ is not the trivial map. By Theorem 2.3.27, $\mathrm{Aut}(P) \cong U_{19}$ which is an abelian group of order $\phi(19) = 18$. Since Q has order 9, this implies $\ker(\theta)$ has order 1 or 3, and $\mathrm{im}(\theta)$ has order 3 or 9. A direct computation shows that U_{19} is cyclic and has 6 elements of order 9, namely [4], [5], [6], [9], [16], and [17]. The 2 elements of order 3 are [7] and [11]. There are three cases.

Case 1: Assume $Q = \langle b \rangle$ is cyclic and $\operatorname{im}\theta$ has order 9. Then θ maps Q isomorphically onto the subgroup of order 9 in $\operatorname{Aut}(P)$. If necessary, we replace b with the generator of Q that maps to $[4] \in U_{19}$. We have $bab^{-1} = a^4$. The presentation of G in terms of generators and relations is

$$G = \langle a, b \mid a^{19} = b^9 = e,\, bab^{-1} = a^4 \rangle.$$

Case 2: Assume $Q = \langle b \rangle$ is cyclic and $\operatorname{im}\theta$ has order 3. Then the kernel of θ is the cyclic subgroup of order 3. Under θ, an element of order 9 is mapped onto one of the elements of order 3. If necessary, we replace b with a generator of Q that maps to $[7] \in U_{19}$. We have $bab^{-1} = a^7$. The presentation of G in terms of generators and relations is

$$G = \langle a, b \mid a^{19} = b^9 = e,\, bab^{-1} = a^7 \rangle.$$

Case 3: Assume Q is a direct sum of two cyclic groups of order 3. Since U_{19} has a unique subgroup of order 3, the kernel of θ is a group of order 3. Suppose $\ker(\theta) = \langle c \rangle$. Because the image of θ contains both $[7]$ and $[11]$, we pick $b \in Q - \langle c \rangle$ such that $\theta(b) = [7]$. Then $Q = \langle b, c \rangle$, $cac^{-1} = a$, and $bab^{-1} = a^7$. Then in this case,

$$G = \langle a, b, c \mid a^{19} = b^3 = c^3 = e,\, bc = cb,\, bab^{-1} = a^7,\, cac^{-1} = a \rangle$$

is the presentation of G.

9.5. Groups of order 225. In this example we show that there are at least six nonisomorphic groups of order 225. We show how to construct two nonisomorphic nonabelian groups of order $225 = 3^2 5^2$. Let G denote a group of order 225. Let P be a 5-Sylow subgroup of G. By Theorem 2.7.7, the number of conjugates of P divides 9 and is congruent to 1 modulo 5. We conclude that P is normal in G. Let Q be a 3-Sylow subgroup of G. The number of conjugates of Q divides 25 and is congruent to 1 modulo 3. Therefore, either Q is normal in G, or Q has 25 conjugates. By Theorem 2.7.1 (2), both P and Q are abelian.

Case 1: Assume P and Q are both normal in G. By Theorem 2.7.8, G is the internal direct product of P and Q, hence G is abelian. By Theorem 2.8.7, G is isomorphic to either

$$\mathbb{Z}/9 \times \mathbb{Z}/25$$

or

$$\mathbb{Z}/9 \times \mathbb{Z}/5 \times \mathbb{Z}/5$$

or

$$\mathbb{Z}/3 \times \mathbb{Z}/3 \times \mathbb{Z}/25$$

or

$$\mathbb{Z}/3 \times \mathbb{Z}/3 \times \mathbb{Z}/5 \times \mathbb{Z}/5.$$

Case 2: Assume P is normal and Q has 25 conjugates. Then Q acts by conjugation on P and there is a homomorphism of groups $\theta : Q \to \operatorname{Aut}(P)$. There are two subcases to consider.

Subcase 2.1: Assume P is cyclic. By Theorem 2.3.27, $\operatorname{Aut}(P) \cong U_{25}$ is an abelian group of order $\phi(25) = 20$. Since $\operatorname{Aut}(P)$ has no subgroup of order 3, θ is the trivial homomorphism. Therefore, every element of Q commutes with every element of P. By Exercise 2.5.19, G is the internal direct product of P and Q, hence this case reduces to Case 1.

Subcase 2.2: Assume $P \cong \mathbb{Z}/5 \times \mathbb{Z}/5$. Then $\operatorname{Aut}(P)$ is isomorphic to $\mathrm{GL}_2(\mathbb{Z}/5)$. We will prove this in Proposition 4.5.8. As seen in Exercise 2.9.5, there are subgroups of order 3 in $\operatorname{Aut}(P)$. Without being more specific, we end this example by showing how to construct two nonisomorphic nonabelian groups of order 225. Let $\alpha \in \operatorname{Aut}(P)$ be an automorphism of P of order 3. There are two cases for Q.

Subcase 2.2.1: Assume $Q = \langle a \mid a^9 = e\rangle$ is cyclic of order 9. Then $a \mapsto \alpha$ induces $\theta : Q \to \mathrm{Aut}(P)$. The kernel of θ has order 3, the image of θ has order 3. Then the semidirect product $P \rtimes Q$ is a nonabelian group of order 225.

Subcase 2.2.2: Assume $Q = \langle a, b \mid a^3 = b^3 = e\rangle$ is a noncyclic group of order 9. Then $a \mapsto \alpha$, $b \mapsto e$ induces $\theta : Q \to \mathrm{Aut}(P)$. The kernel of θ is $\langle b\rangle$, which has order 3, the image of θ is $\langle\alpha\rangle$, which has order 3. Then the semidirect product $P \rtimes Q$ is a nonabelian group of order 225.

9.6. Groups of order p^3. Let p be an odd prime. In this example we show how to construct a nonabelian group of order p^3. Let F be the field $\mathbb{Z}/p$. Let $V = F^2 = \{(x_1, x_2) \mid x_i \in F\}$ where the binary operation on V is written additively. Then V is isomorphic to $\mathbb{Z}/p \times \mathbb{Z}/p$. Let $\theta \in \mathrm{GL}_2(F)$ be the matrix $\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$. Then $\theta^2 = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$, $\theta^3 = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$, $\ldots$, $\theta^{p-1} = \begin{bmatrix} 1 & 0 \\ p-1 & 1 \end{bmatrix}$, $\theta^p = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. This shows that $C = \langle\theta\rangle$ is a cyclic subgroup of $\mathrm{GL}_2(F)$ of order p. Although we have not proved it yet, using matrices and properties of Hom we will prove in Proposition 4.5.8 that $\mathrm{Aut}(V) \cong \mathrm{GL}_2(F)$. Therefore, the semidirect product $V \rtimes C$ is a nonabelian group of order p^3 containing a normal subgroup isomorphic to V. Before ending this example, we show that every element of the semidirect product has order 1 or p. Let $i \in \mathbb{Z}$. Then

$$\begin{aligned} I_2 + \theta^i + \theta^{2i} + \cdots + \theta^{(p-1)i} &= \begin{bmatrix} p & 0 \\ 0 + i + 2i + \cdots + (p-1)i & p \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ ip(p-1)/2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Let $z = (x, \theta^i)$ be a typical element of the semidirect product $V \rtimes C$. Then

$$\begin{aligned} z^2 &= (x, \theta^i)(x, \theta^i) = (x + \theta^i(x), \theta^{2i}) = \left((I_2 + \theta^i)(x), \theta^{2i}\right) \\ z^3 &= \left((I_2 + \theta^i)(x), \theta^{2i}\right)(x, \theta^i) = \left((I_2 + \theta^i + \theta^{2i})(x), \theta^{3i}\right) \\ &\vdots \\ z^p &= \left((I_2 + \theta^i + \theta^{2i} + \cdots + \theta^{(p-1)i})(x), \theta^{pi}\right) = (0, I_2). \end{aligned}$$

This shows z has order 1 or p.

9.7. Exercises.

EXERCISE 2.9.1. This exercise is a continuation of Example 9.3. Let G be a nonabelian group of order 63. Show that G contains a cyclic subgroup N of order 21 and N is normal in G. Show that the center of G is a cyclic group of order 3.

EXERCISE 2.9.2. Classify up to isomorphism all groups of order 99.

EXERCISE 2.9.3. Show that up to isomorphism there are 5 groups of order 8, namely $\mathbb{Z}/8$, $\mathbb{Z}/4 \times \mathbb{Z}/2$, $\mathbb{Z}/2 \times \mathbb{Z}/2 \times \mathbb{Z}/2$, the dihedral group D_4, and the quaternion 8-group Q_8 of Example 2.1.18.

EXERCISE 2.9.4. (The square roots of unity in $\mathrm{GL}_2(\mathbb{Z}/5)$) The general linear group of 2-by-2 matrices over the field $\mathbb{Z}/5$, denoted $\mathrm{GL}_2(\mathbb{Z}/5)$, is the multiplicative group of

invertible matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with entries in the field $\mathbb{Z}/5$. In this exercise the reader is asked to find all matrices M in $\mathrm{GL}_2(\mathbb{Z}/5)$, such that $M^2 = I_2$, where I_2 denotes the identity matrix. The following is a suggested outline to show that there are 31 elements of order two in $\mathrm{GL}_2(\mathbb{Z}/5)$.

(1) Let $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and assume $M^2 = I_2$. Show that a, b, c, d satisfy the equations: $a^2 - d^2 = 0$, $bc = 1 - a^2$.
(2) If $a = 0$, then M is of the form $\begin{bmatrix} 0 & b \\ b^{-1} & 0 \end{bmatrix}$, where $b = 1, 2, 3, 4$, so there are 4 such matrices.
(3) If $a = \pm 1$, then M has one of the forms $\pm \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\pm \begin{bmatrix} 1 & b \\ 0 & -1 \end{bmatrix}$, $\pm \begin{bmatrix} 1 & 0 \\ c & -1 \end{bmatrix}$, where $b = 0, 1, 2, 3, 4$, $c = 1, 2, 3, 4$. There are 20 such matrices, one of them has order 1, the rest order 2.
(4) If $a = \pm 2$, then M has one of the forms $\pm \begin{bmatrix} 2 & b \\ c & -2 \end{bmatrix}$, where $bc = 2$. There are 8 such matrices.

EXERCISE 2.9.5. (The cube roots of unity in $\mathrm{GL}_2(\mathbb{Z}/5)$) The general linear group of 2-by-2 matrices over the field $\mathbb{Z}/5$, denoted $\mathrm{GL}_2(\mathbb{Z}/5)$, is the multiplicative group of invertible matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with entries in the field $\mathbb{Z}/5$. In this exercise the reader is asked to find all matrices M in $\mathrm{GL}_2(\mathbb{Z}/5)$, such that $M^3 = I_2$, where I_2 denotes the identity matrix. The following is a three-step outline to show that there are 20 elements of order three in $\mathrm{GL}_2(\mathbb{Z}/5)$.

(1) Let $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Show that if $M^2 + M + I_2 = 0$, then $M^3 = I_2$.
(2) Show that a, b, c, d satisfy the equations: $bc = -(a^2 + a + 1)$, $d = 4 - a$.
(3) Show that there are 5 choices for a and for each a there are 4 choices for the ordered triple (b, c, d).
(4) This part assumes the reader has basic familiarity with field extensions. Show that every element of order three in the ring of 2-by-2 matrices over the field $\mathbb{Z}/5$ is a root of the polynomial equation $x^2 + x + 1 = 0$. Prove that every element of order 3 in $\mathrm{GL}_2(\mathbb{Z}/5)$ is in the list of Part (3).

10. Chain Conditions

10.1. Nilpotent Groups and Solvable Groups.

DEFINITION 2.10.1. Let G be a group. Set $Z^0 = \langle e \rangle$ and $Z^1 = Z(G)$, the center of G. Then $Z^1 = \{x \in G \mid xyx^{-1}y^{-1} \in Z^0 \text{ for all } y \in G\}$. By Exercise 2.3.38, Z^1 is an abelian normal subgroup of G. Inductively assume that $n \geq 1$ and we have the chain of normal subgroups $Z^0 \subseteq Z^1 \subseteq \cdots \subseteq Z^n$ in G. Let $\eta_n : G \to G/Z^n$ be the natural map. Then Z^{n+1} is defined by the rules

$$\begin{aligned} Z^{n+1} &= \eta_n^{-1}(Z(G/Z^n)) \\ &= \{x \in G \mid xyx^{-1}y^{-1} \in Z^n \text{ for all } y \in G\}. \end{aligned}$$

By Theorem 2.3.13, Z^{n+1} is a normal subgroup of G, $Z^n \subseteq Z^{n+1}$, and the quotient group Z^{n+1}/Z^n is isomorphic to $Z(G/Z^n)$, hence is abelian. The ascending chain of subgroups $Z^0 \subseteq Z^1 \subseteq Z^2 \subseteq \cdots \subseteq Z^n \subseteq Z^{n+1} \subseteq \cdots$ is called the *ascending central series of G*.

DEFINITION 2.10.2. Let G be a group. We say G is *nilpotent*, if the ascending central series of G converges to G. That is, if $Z^n = G$ for some $n \geq 1$.

LEMMA 2.10.3. *Let p be a prime and G a finite p-group. Then G is nilpotent.*

PROOF. By Theorem 2.7.1, G has a nontrivial center. If G is abelian, then $Z^1 = G$. Otherwise, $Z^1 \subsetneq G$, and the quotient G/Z^1 is a p-group of order less than $|G|$. Since G is finite, $Z^n = G$ for some $n \geq 1$. □

LEMMA 2.10.4. *If A and B are groups, then $Z^n(A \times B) = Z^n(A) \times Z^n(B)$.*

PROOF. The proof is by induction on n. By Exercise 2.3.38, $Z(A \times B) = Z(A) \times Z(B)$, so the result is true for $n = 1$. Assume inductively that $j \geq 1$ and $Z^j(A \times B) = Z^j(A) \times Z^j(B)$. By Exercise 2.5.20,

$$\frac{A \times B}{Z^j(A \times B)} = \frac{A \times B}{Z^j(A) \times Z^j(B)} = \frac{A}{Z^j(A)} \times \frac{B}{Z^j(B)}.$$

By Exercises 2.3.38 and 2.5.20,

$$\begin{aligned} Z\left(\frac{A \times B}{Z^j(A \times B)}\right) &= Z\left(\frac{A}{Z^j(A)} \times \frac{B}{Z^j(B)}\right) \\ &= Z\left(\frac{A}{Z^j(A)}\right) \times Z\left(\frac{B}{Z^j(B)}\right) \\ &= \frac{Z^{j+1}(A)}{Z^j(A)} \times \frac{Z^{j+1}(B)}{Z^j(B)} \\ &= \frac{Z^{j+1}(A) \times Z^{j+1}(B)}{Z^j(A) \times Z^j(B)} \\ &= \frac{Z^{j+1}(A) \times Z^{j+1}(B)}{Z^j(A \times B)}. \end{aligned}$$

This proves $Z^{j+1}(A \times B)/Z^j(A \times B) = \left(Z^{j+1}(A) \times Z^{j+1}(B)\right)/Z^j(A \times B)$. It follows from Theorem 2.3.13 that $Z^{j+1}(A \times B) = Z^{j+1}(A) \times Z^{j+1}(B)$. This completes the proof. □

PROPOSITION 2.10.5. *The direct product of a finite number of nilpotent groups is nilpotent.*

PROOF. Let A and B be nilpotent groups. We show that $A \times B$ is nilpotent. A finite induction argument proves the result for a general finite product. By hypothesis, there exists $n \geq 1$ such that $A = Z^m(A)$ and $B = Z^m(B)$. By Lemma $Z^m(A \times B) = Z^m(A) \times Z^m(B) = A \times B$. □

LEMMA 2.10.6. *Let G be a nilpotent group and H a proper subgroup of G. Then H is a proper subgroup of $N_G(H)$, the normalizer of H in G.*

PROOF. For some $n \geq 1$, we are given that $Z^n = G$. Let k be the largest integer such that $Z^k \subseteq H$. Let $a \in Z^{k+1} - H$. Then $aha^{-1} \equiv h \pmod{Z^k}$ implies there exists $z \in Z^k$ such that $aha^{-1} = zh$. But $zh \in H$, hence $a \in N_G(H) - H$. □

THEOREM 2.10.7. *Let G be a finite group. Then G is nilpotent if and only if G is the internal direct product of its Sylow subgroups.*

PROOF. Assume G is a finite nilpotent group. Let p be a prime divisor of $|G|$ and P a Sylow p-subgroup of G. First we show that P is a normal subgroup of G. By Corollary 2.7.6 (3), $N_G(N_G(P)) = N_G(P)$. By Lemma 2.10.6, $N_G(P) = G$. By Proposition 2.4.12, P is a normal subgroup of $N_G(P) = G$. By Proposition 2.7.8, G is the internal direct product of its Sylow subgroups. The converse follows from Lemma 2.10.3 and Proposition 2.10.5. □

DEFINITION 2.10.8. Let G be a group. By Exercise 2.3.42, the commutator subgroup of G, denoted G', is the subgroup of G generated by the set $\{xyx^{-1}y^{-1} \mid x, y \in G\}$. Moreover, G' is a normal subgroup of G and the quotient group G/G' is abelian. Set $G^{(0)} = G$ and $G^{(1)} = G'$. Recursively, for $n \geq 1$, define G^{n+1} to be the commutator subgroup of $G^{(n)}$. Then G^{n+1} is a normal subgroup of $G^{(n)}$ and the quotient group $G^{(n)}/G^{(n+1)}$ is an abelian group. The descending chain of subgroups $G^{(0)} \supseteq G^{(1)} \supseteq G^{(2)} \supseteq \cdots \supseteq G^{(n)} \supseteq G^{(n+1)} \supseteq \cdots \supseteq \langle e \rangle$ is called the *derived series of G*.

DEFINITION 2.10.9. A group G is said to be *solvable* if there is a descending chain of subgroups $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_m = \langle e \rangle$ starting with G and ending with $\langle e \rangle$ such that for $0 < i \leq m$, G_i is a normal subgroup of G_{i-1} and the quotient G_i/G_{i-1} is an abelian group. In this case, we say $G_0, G_1, \ldots, G_m$ is a *solvable series* for G.

EXAMPLE 2.10.10. It is proved in Theorem 2.7.1 that a finite p-group is solvable.

EXAMPLE 2.10.11. If G is a finite abelian group, then $\langle e \rangle \subseteq G$ is a solvable series for G.

LEMMA 2.10.12. *Let G be a group. If G is nilpotent, that is, if there exists $k \geq 1$ such that $Z^k = G$, then G is solvable.*

PROOF. Assume the ascending central series $\langle e \rangle = Z^0 \subseteq Z^1 \subseteq Z^2 \subseteq \cdots \subseteq Z^{k-1} \subseteq Z^k = G$ begins at $\langle e \rangle$ and ends at G. Since each quotient Z^{n+1}/Z^n is abelian, this is a solvable series. □

LEMMA 2.10.13. *Let G be a group. Then G has a solvable series if and only if for some $k \geq 1$, the kth derived subgroup $G^{(k)}$ is equal to $\langle e \rangle$. In other words, G is solvable if and only if the derived series converges to $\langle e \rangle$.*

PROOF. If $G^{(k)} = \langle e \rangle$, then the derived series is a solvable series. Conversely, assume $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_m = \langle e \rangle$. Since G_1 is a normal subgroup of G and G/G_1 is abelian, by Exercise 2.3.42 (3), $G' \subseteq G_1$. Then $\{aba^{-1}b^{-1} \mid a, b \in G'\}$ is a subset of $\{aba^{-1}b^{-1} \mid a, b \in G_1\}$. So $G^{(2)} = G'' \subseteq G_1'$. But G_2 is a normal subgroup of G_1 and G_1/G_2 is abelian, so $G_1' \subseteq G_2$. Taken together, we have $G^{(2)} \subseteq G_2$. Iterating this argument shows that $G^{(m)} \subseteq G_m = \langle e \rangle$. □

COROLLARY 2.10.14. *The symmetric group S_n is solvable if and only if $n \leq 4$.*

PROOF. A solvable series for S_3 is $\langle e \rangle \subseteq A_3 = \langle e, (123), (132) \rangle \subseteq S_3$. A solvable series for S_4 is $\langle e \rangle \subseteq \langle e, (12)(34), (13)(24), (14)(23) \rangle \subseteq A_4 \subseteq S_4$. Let $n \geq 5$ and let $G = S_n$. Since S_n/A_n is cyclic of order two, by Exercise 2.3.42 (3), $G' \subseteq A_n$. Since A_n is nonabelian and simple, $G' = G^{(2)} = A_n$. Therefore, the derived series for G converges to A_n. By Lemma 2.10.13, G is not solvable. □

10.2. Composition Series.

DEFINITION 2.10.15. Let G be a group and suppose there is a strictly descending finite chain of subgroups

$$G = G_0 \supsetneq G_1 \supsetneq G_2 \supsetneq \cdots \supsetneq G_n = \langle e \rangle$$

starting with $G = G_0$ and ending with $G_n = \langle e \rangle$. The *length* of the chain is n. A *composition series* for G is a chain such that for $i = 1, \ldots, n$, G_i is a normal subgroup of G_{i-1} and G_{i-1}/G_i is simple. If G has no composition series, define $\ell(G) = \infty$. Otherwise, let $\ell(G)$ be the minimum of the lengths of all composition series of G.

LEMMA 2.10.16. *Let G be a finite group. Then G has a composition series.*

PROOF. The reader should verify that a strictly descending chain of subgroups of maximum length such that G_i is a normal subgroup of G_{i-1} is a composition series. □

10.3. Exercises.

EXERCISE 2.10.17. Let G be a group. Prove:

(1) For each $k \geq 1$, the kth derived subgroup, $G^{(k)}$, is a normal subgroup of G.
(2) If $\theta : G \to H$ is an epimorphism, then $\theta(G^{(k)}) = H^{(k)}$.

EXERCISE 2.10.18. Let G be a group. Prove:

(1) If G is solvable and H is a subgroup of G, then H is solvable.
(2) If G is solvable and $\theta : G \to H$ is an epimorphism, then H is solvable.
(3) Let N be a normal subgroup of G. If N and G/N are solvable, then G is solvable.
(4) If $G \neq \langle e \rangle$ and G is solvable, then there exists an abelian normal subgroup $A \subseteq G$, $A \neq \langle e \rangle$.

EXERCISE 2.10.19. Let $n \geq 3$.

(1) Show that there is a homomorphism $\theta : D_{2n} \to D_n$ from the dihedral group D_{2n} onto the dihedral group D_n and the kernel of θ is the center of D_{2n}. (Hint: Example 2.3.32.)
(2) Let 2^m be the highest power of 2 that divides n. Show that the central ascending series of D_n is $Z^{(0)} \subseteq Z^{(1)} \subseteq \cdots \subseteq Z^{(m)}$, where $Z^{(i)} = \langle R^{n/2^i} \rangle$.
(3) Show that if n is odd, then D_{2n} is the internal direct sum of a cyclic subgroup of order two (the center) and a subgroup isomorphic to D_n.

EXERCISE 2.10.20. Let G be a finite solvable group. Prove:

(1) If G is abelian and $G = G_0 \supsetneq G_1 \supsetneq \cdots \supsetneq G_m = \langle e \rangle$ is a composition series, then G_{i-1}/G_i is a cyclic group and $[G_{i-1} : G_i]$ is a prime number.
(2) G has a composition series $G = G_0 \supsetneq G_1 \supsetneq \cdots \supsetneq G_m = \langle e \rangle$ such that G_{i-1}/G_i is a cyclic group and $[G_{i-1} : G_i]$ is a prime number.

CHAPTER 3

Rings

A ring is an algebraic structure which has two binary operations called addition and multiplication. We have already seen concrete examples of rings. The prototypical example of a ring is the ring of integers, $\mathbb{Z}$. Its close relative is the ring of integers modulo n, $\mathbb{Z}/(n)$. The fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are rings. The ring of n-by-n matrices $M_n(\mathbb{R})$ is an example of a ring in which multiplication is not commutative. The set of polynomials, the set of rational functions, and the set of power series with coefficients over the field $\mathbb{R}$ are rings. The set of all continuous functions, differentiable functions, and integrable functions from $\mathbb{R}$ to $\mathbb{R}$ are rings. The set of all functions from $\mathbb{R}$ to $\mathbb{R}$ that are continuous at a specific point is a ring. If A is an abelian group, the set of all endomorphisms from A to itself is a ring. Ring Theory can be viewed as the axiomatic abstraction of these examples.

1. Definitions and Terminology

DEFINITION 3.1.1. A *ring* is a nonempty set R with two binary operations, addition written $+$, and multiplication written $\cdot$ or by juxtaposition. Under addition $(R,+)$ is an abelian group with identity element 0. Under multiplication $(R,\cdot)$ is associative and contains an identity element, denoted by 1. Multiplication distributes over addition from both the left and the right. If $(R,\cdot)$ is commutative, then we say R is a *commutative ring*. The *trivial ring* is $\{0\}$, in which $0 = 1$. If R is not the trivial ring, the reader is asked to prove in Proposition 3.1.2 that $0 \neq 1$.

PROPOSITION 3.1.2. *Let R be a ring. Let $a,b \in R$, $n,m \in \mathbb{N}$, $a_1,\ldots,a_n,b_1,\ldots,b_m \in R$.*

(1) $0a = a0 = 0$.
(2) $(-a)b = a(-b) = -(ab)$.
(3) $(-a)(-b) = ab$.
(4) $(na)b = a(nb) = n(ab)$.
(5) $\left(\sum_{i=1}^{n} a_i\right)\left(\sum_{j=1}^{m} b_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j$
(6) If R contains more than one element, then $0 \neq 1$.

PROOF. Is left to the reader. □

DEFINITION 3.1.3. Let R be a ring and $a \in R$. We say a is a *left zero divisor* if $a \neq 0$ and there exists $b \neq 0$ such that $ab = 0$. We say a is *left invertible* in case there is $b \in R$ such that $ba = 1$. The reader should define the terms *right zero divisor* and *right invertible*. If a is both a left zero divisor and right zero divisor, then we say a is a *zero divisor*. If a is both left invertible and right invertible, then we say a is *invertible*. In this case, the left inverse and right inverse of a are equal and unique (Exercise 2.1.22 (2)). An invertible element in a ring R is also called a *unit* of R. If $R \neq (0)$ and R has no zero divisors, then we say R is a *domain*. A commutative domain is called an *integral domain*. A domain in which every nonzero element is invertible is called a *division ring*. A commutative division ring

is called a *field*. The set of all invertible elements in a ring R is a group which is denoted $\text{Units}(R)$ or R^* and is called *the group of units in R*.

REMARK 3.1.4. Notice that in Definition 3.1.3, we have explicitly required a domain to have at least two elements. The only ring with order one is the trivial ring (0). In Example 3.2.4 (4) we see that (0) plays the role of a terminal object in the category of rings. Besides this, there is no significant result that can be proved about the ring (0). It has no proper ideals, is not a subring of any larger ring, and there is no nontrivial module or algebra over (0).

EXAMPLE 3.1.5. Standard examples of rings and fields are listed here.

(1) The ring of integers $\mathbb{Z}$ is an integral domain. The ring of integers modulo n, denoted $\mathbb{Z}/(n)$, is a commutative ring containing n elements.
(2) Denote by $\mathbb{Q}$ the field of rational numbers, by $\mathbb{R}$ the field of real numbers and by $\mathbb{C}$ the field of complex numbers (see Section 1.4).
(3) If k is a field and $n \geq 1$, the ring of n-by-n matrices over k is denoted by $M_n(k)$. If $n > 1$, then $M_n(k)$ is noncommutative.
(4) If R is any ring, the ring of n-by-n matrices over R is denoted by $M_n(R)$.

EXAMPLE 3.1.6. Let R be a commutative ring and G a finite multiplicative group. Assume the order of G is n and enumerate the elements $G = \{g_1, \ldots, g_n\}$, starting with the group identity $g_1 = e$. Let $R(G)$ be the set of all formal sums

$$R(G) = \{r_1 g_1 + \cdots + r_n g_n \mid r_i \in R\}.$$

Define addition and multiplication rules on $R(G)$ by

$$\sum_{i=1}^{n} r_i g_i + \sum_{i=1}^{n} s_i g_i = \sum_{i=1}^{n} (r_i + s_i) g_i$$
$$\left(\sum_{i=1}^{n} r_i g_i\right)\left(\sum_{i=1}^{n} s_i g_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} (r_i s_j)(g_i g_j)$$

The additive identity is $0 = 0g_1 + 0g_2 + \cdots + 0g_n$. The multiplicative identity is $1 = 1g_1 + 0g_2 + \cdots + 0g_n$. Then $R(G)$ is a ring. We call $R(G)$ a *group ring*.

If R is a commutative ring and G is a group which is not necessarily finite, we can still define the group ring $R(G)$. In this case, take $R(G)$ to be the set of all finite formal sums

$$R(G) = \left\{ \sum_{g \in G} r_g g \mid r_g \in R \text{ and } r_g = 0 \text{ for all but finitely many } g \right\}.$$

If $g \in G$, then in $R(G)$ we have the identity $gg^{-1} = g^{-1}g = 1$. Therefore, we can view G as a subgroup of the group of units in the group ring $R(G)$.

EXAMPLE 3.1.7. If A is an abelian group, let $\text{Hom}(A,A)$ be the set of all homomorphisms from A to A. Turn $\text{Hom}(A,A)$ into a ring by coordinate-wise addition and composition of functions:

$$(f+g)(x) = f(x) + g(x)$$
$$(fg)(x) = f(g(x))$$

See Exercise 2.8.11. For computations of $\text{Hom}(\mathbb{Z},\mathbb{Z}) \cong \mathbb{Z}$ and $\text{Hom}(\mathbb{Z}/n,\mathbb{Z}/n) \cong \mathbb{Z}/n$, see Exercises 3.1.16 and 3.1.17.

DEFINITION 3.1.8. If R is any ring, the *opposite ring of* R is denoted R^o. As an additive abelian group, the opposite ring of R is equal to R. However, the multiplication of R^o is reversed from that of R. Writing the multiplication of R by juxtaposition and multiplication of R^o with the asterisk symbol, we have $x * y = yx$.

DEFINITION 3.1.9. If A is a ring and $B \subseteq A$, then we say B is a *subring* of A if B contains both 0 and 1 and B is a ring under the addition and multiplication rules of A. Let A be a ring. The *center* of A is the set

$$Z(A) = \{x \in A \mid xy = yx\,(\forall y \in A)\}.$$

The reader should verify that $Z(A)$ is a subring of A and $Z(A)$ is a commutative ring. If $x \in Z(R)$, then we say x is *central*.

EXAMPLE 3.1.10. Let $R = \mathbb{Z}/6 = \{0,1,2,3,4,5\}$ be the ring of integers modulo 6. Let $B = \{0,2,4\}$ and $C = \{0,3\}$. The reader should verify that B is a ring of order 3. In fact, B is isomorphic to the field $\mathbb{Z}/3$. Since B does not contain 1, B is not a subring of R. Likewise, C is a ring, isomorphic to the field $\mathbb{Z}/2$, but C is not a subring of R. The sets B and C are examples of ideals (see Example 3.2.2).

EXAMPLE 3.1.11. If $n > 1$, then the additive group $(\mathbb{Z}/n, +)$ is generated by 1. Therefore, the ring $\mathbb{Z}/n$ has no proper subring.

EXAMPLE 3.1.12. Let R be any ring and $M_n(R)$ the ring of n-by-n matrices over R, where $n \geq 2$. The set

$$L = \left\{(r_{ij}) \mid r_{ij} = 0 \text{ if } i < j\right\}$$

of all lower triangular matrices is a noncommutative subring of $M_n(R)$. Likewise, the set of all upper triangular matrices is a noncommutative subring of $M_n(R)$. See Example 3.3.10 for a continuation of this example when R is a field and $n = 2$.

EXAMPLE 3.1.13. Let R be a commutative ring and $M_2(R)$ the ring of two-by-two matrices over R. The proof given in Example 2.3.34 can be readily adapted to show that the center of the ring $M_2(R)$ is equal to the set of scalar matrices $\left\{\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \mid a \in R\right\}$. Let $n \geq 2$. Using a different proof, we show that the center of the ring $M_n(R)$ is equal to the set of scalar matrices over R. Let $A = (a_{ij})$ be a central matrix. For each ordered pair (i, j), where $1 \leq i, j \leq n$, let e_{ij} be the elementary matrix with 1 in position (i, j) and 0 elsewhere. In the following, we use the following notation: $C_i(A)$ denotes column i of A, $R_j(A)$ denotes row j of A, and $M_{rs}(0)$ denotes the r-by-s matrix with 0 in every position. Then

$$e_{ij}A = \begin{pmatrix} M_{i-1,n}(0) \\ R_j(A) \\ M_{n-i,n}(0) \end{pmatrix}.$$

In words, row i of $e_{ij}A$ is equal to row j of A and all other entries of $e_{ij}A$ are equal to 0. Also,

$$Ae_{ij} = \begin{pmatrix} M_{n,j-1}(0) & C_i(A) & M_{n,n-j}(0) \end{pmatrix}.$$

In words, column j of Ae_{ij} is equal to column i of A and all other entries of $e_{ij}A$ are equal to 0. Since A commutes with e_{ij}, we conclude that all elements of A that are not on the diagonal are equal to 0. If we assume $i \neq j$, this also means $a_{jj} = a_{ii}$. Therefore, A is a scalar matrix. It is routine to check that a scalar matrix is central.

EXAMPLE 3.1.14. If F is a field the ring of quaternions over F is the four-dimensional vector space over F with basis $\{1,i,j,k\}$ with multiplication defined by extending these relations:

$$\begin{aligned} i^2 = j^2 = k^2 &= -1 \\ ij &= -ji = k \\ ik &= -ki = -j \end{aligned}$$

by associativity and distributivity. We denote the ring of quaternions by $\mathbb{H}(F)$, or $\mathbb{H}_F$. Notice that under multiplication the set $\{1,-1,i,-i,j,-j,k,-k\}$ is Q_8, the quaternion 8-group of Example 2.1.18. The ring of quaternions $\mathbb{H}_F$ is a division ring if F is equal to either $\mathbb{Q}$ or $\mathbb{R}$ (Exercise 3.1.18). The ring of quaternions $\mathbb{H}_{\mathbb{C}}$ is isomorphic to $M_2(\mathbb{C})$ (Exercise 3.1.20). The ring of quaternions $\mathbb{H}(\mathbb{Z}/(2))$ is commutative (Exercise 3.1.19). The product formula for multiplying two quaternions $x = a+bi+cj+dk$ and $y = e+fi+gj+hk$ is

$$\begin{aligned} xy &= (a+bi+cj+dk)(e+fi+gj+hk) \\ &= (ae-bf-cg-dh)+(af+be+ch-dg)i \\ &\quad +(ag-bh+ce+df)j+(ah+bg-cf+de)k \end{aligned}$$

and is derived from the relations above. We identify F with $F\cdot 1$. Thus, F is a subring of $\mathbb{H}_F$. If $x \in F$, then $xy = yx$. That is, F is a subring of the center of $\mathbb{H}_F$. For a quaternion $x = a+bi+cj+dk$ define $\chi(x) = a-bi-cj-dk$. Using the product formula above, we find

$$\begin{aligned} \chi(y)\chi(x) &= (e-fi-gj-hk)(a-bi-cj-dk) \\ &= (ae-bf-cg-dh)-(af+be+ch-dg)i \\ &\quad -(ag-bh+ce+df)j-(ah+bg-cf+de)k \\ &= \chi(xy). \end{aligned}$$

Define the *norm* of x by

$$\begin{aligned} N(x) = x\chi(x) &= (a+bi+cj+dk)(a-bi-cj-dk) \\ &= (a^2+b^2+c^2+d^2)+(-ab+ab+cd-cd)i \\ &\quad +(ac+bd-ac-bd)j+(-ad-bc+bc+ad)k \\ &= a^2+b^2+c^2+d^2 \end{aligned}$$

which is an element of F. Using the formulas from above, we see that

$$N(xy) = xy\chi(xy) = xy\chi(y)\chi(x) = xN(y)\chi(x) = x\chi(x)N(y) = N(x)N(y)$$

hence $N : \mathbb{H}_F \to F$ is multiplicative. The function χ is an example of an involution.

DEFINITION 3.1.15. Let R and S be rings. A function $\theta : R \to S$ is called an *isomorphism of rings*, if θ is a one-to-one correspondence, $\theta(1) = 1$, $\theta(x+y) = \theta(x)+\theta(y)$, and $\theta(xy) = \theta(x)\theta(y)$ for all $x,y \in R$. In this case, we say R and S are *isomorphic* and write $R \cong S$. From an abstract algebraic point of view, isomorphic rings are indistinguishable.

1.1. Exercises.

EXERCISE 3.1.16. The point to this exercise is to compute the ring $\mathrm{Hom}(\mathbb{Z},\mathbb{Z})$ of all endomorphisms of the infinite cyclic group $(\mathbb{Z},+)$ (see Exercise 2.8.11). In the following, f and g always denote endomorphisms of $\mathbb{Z}$.

(1) Define $\phi : \mathrm{Hom}((\mathbb{Z},+),(\mathbb{Z},+)) \to \mathbb{Z}$ by $\phi(f) = f(1)$. Show that ϕ is an isomorphism of rings. (Hint: Theorem 2.3.27.)
(2) Show that $\mathrm{Aut}((\mathbb{Z},+))$ has order two.

EXERCISE 3.1.17. Let $n \in \mathbb{N}$. The object of this exercise is to compute the ring of all endomorphisms of the finite cyclic group $(\mathbb{Z}/n,+)$. As in Exercise 2.8.11, this ring is denoted $\mathrm{Hom}((\mathbb{Z}/n,+),(\mathbb{Z}/n,+))$. In the following, f and g always denote endomorphisms of $(\mathbb{Z}/n,+)$.

(1) Define $\phi : \mathrm{Hom}((\mathbb{Z}/n,+),(\mathbb{Z}/n,+)) \to \mathbb{Z}/n$ by $\phi(f) = f(1)$. Prove that ϕ is an isomorphism of rings. (Hint: Theorem 2.3.27.)
(2) Show that $\mathrm{Aut}((\mathbb{Z}/n,+)) \cong U_n$, where U_n is the group of units modulo n.

EXERCISE 3.1.18. Prove that the ring of quaternions (see Example 3.1.14) over $\mathbb{Q}$ (or $\mathbb{R}$) is a division ring.

EXERCISE 3.1.19. Let $G = \langle a,b \mid a^2 = b^2 = e,\ ab = ba \rangle$ be an elementary 2-group of order 4. Let $R = \mathbb{Z}/(2)$ be the field with 2 elements. For the definition of the ring of quaternions, see Example 3.1.14. For the definition of a group ring, see Example 3.1.6.

(1) Prove that the ring of quaternions over R is isomorphic to the group ring $R(G)$.
(2) Determine the group of units in $R(G)$.
(3) Determine the set of zero divisors in $R(G)$.
(4) Determine all elements in $R(G)$ that satisfy the equation $e^2 = e$. These elements are the so-called idempotents.

EXERCISE 3.1.20. Prove that the ring of quaternions over $\mathbb{C}$ is isomorphic to $M_2(\mathbb{C})$. (Hint: Find matrices that play the roles of i and j.)

EXERCISE 3.1.21. Let R be the ring $M_2(\mathbb{Z}/(2))$ of two-by-two matrices over $\mathbb{Z}/(2)$.

(1) Determine the group of units in R.
(2) Determine the set of zero divisors in R.
(3) Determine all elements in R that satisfy the equation $e^2 = e$. These elements are the so-called idempotents in R.
(4) Show that R contains exactly two subrings that are fields. One is the image of the canonical homomorphism $\chi : \mathbb{Z} \to R$ which has order 2, and the other is a field of order 4.

EXERCISE 3.1.22. Let R be any ring. Let x and y be elements of R such that $xy = yx$. Prove the Binomial Theorem:

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

for any $n \geq 0$.

EXERCISE 3.1.23. Let $i \in \mathbb{C}$ be the square root of -1.

(1) Show that $\mathbb{Q}[i] = \{a+bi \mid a,b \in \mathbb{Q}\}$ is a subfield of $\mathbb{C}$.
(2) Show that $\mathbb{Z}[i] = \{a+bi \mid a,b \in \mathbb{Q}\}$ is a subring of $\mathbb{Q}[i]$. The ring $\mathbb{Z}[i]$ is called the ring of *gaussian integers*.

EXERCISE 3.1.24. Consider the set

$$\mathbb{Z}/4[i] = \{a+bi \mid a,b \in \mathbb{Z}/4\}$$

where $i^2 = -1 \equiv 3 \pmod 4$. Addition and multiplication are defined as in the gaussian integers, where a and b are added and multiplied in the ring $\mathbb{Z}/4$. Show that $\mathbb{Z}/4[i]$ is a

commutative ring of order 16. Show that the group of units in $\mathbb{Z}/4[i]$ is isomorphic to U_{16}, the group of units modulo 16. Show that the rings $\mathbb{Z}/4[i]$ and $\mathbb{Z}/16$ are not isomorphic.

2. Homomorphisms and Ideals

DEFINITION 3.2.1. Let A be a ring. A *left ideal* of A is a nonempty subset $I \subseteq A$ such that $(I,+)$ is a subgroup of $(A,+)$ and $ax \in I$ for all $a \in A$ and all $x \in I$. The reader should define the term *right ideal*. If I is both a left ideal and right ideal, we say I is an *ideal*.

EXAMPLE 3.2.2. Some important examples of ideals are listed here.

(1) If R is a commutative ring, then a left ideal is a two-sided ideal.
(2) In a ring R the trivial ideals are $\{0\}$ and R.
(3) If F is a field, the only ideals are $\{0\}$ and F. This is Exercise 3.2.32.
(4) Let R be a commutative ring and $M_n(R)$ the ring of n-by-n matrices over R, where $n \geq 2$. The set
$$L = \left\{(r_{ij}) \mid r_{ij} = 0 \text{ if } i < j\right\}$$
of all lower triangular matrices is a subring of $M_n(R)$ (Example 3.1.12). It is not an ideal, because the identity matrix I is in L.
(5) Let F be a field and $M_2(F)$ the ring of 2-by-2 matrices over F. Then
$$I = \left\{ \begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix} \mid a,b \in F \right\}$$
is a left ideal in $M_2(F)$, but not a right ideal.
(6) The subgroups of $\mathbb{Z},+$ are the cyclic subgroups $\mathbb{Z}m$, where $m \in \mathbb{Z}$. Any such subgroup is an ideal. So the ideals of $\mathbb{Z}$ are of the form $\mathbb{Z}m$.

DEFINITION 3.2.3. If R and S are rings, a *homomorphism* from R to S is a function $f\colon R \to S$ satisfying

(1) $f(x+y) = f(x) + f(y)$ for all $x,y \in R$,
(2) $f(xy) = f(x)f(y)$ for all $x,y \in R$, and
(3) $f(1) = 1$.

Notice that (1) implies $f : (R,+) \to (S,+)$ is a homomorphism of additive groups. The *kernel* of f is $\ker(f) = \{x \in R \mid f(x) = 0\}$ which is equal to the kernel of the homomorphism on additive groups. By Exercise 3.2.27, the kernel of f is an ideal in R. By Lemma 2.3.7, f is one-to-one if and only if $\ker f = (0)$. The *image* of the homomorphism f is $\operatorname{im}(f) = \{f(x) \in S \mid x \in R\}$. By Exercise 3.2.27, the image of f is a subring of S. As in Definition 3.1.15, an isomorphism is a homomorphism $f : R \to S$ that is one-to-one and onto. An *automorphism* of R is a homomorphism $f : R \to R$ that is one-to-one and onto.

EXAMPLE 3.2.4. Some important examples of homomorphisms are listed here.

(1) The natural projection $\mathbb{Z} \to \mathbb{Z}/(n)$ maps an integer to its congruence class modulo n. It is a homomorphism of rings which is onto. The kernel is the subgroup generated by n.
(2) If u is an invertible element of R, the *inner automorphism* of R defined by u is $\sigma_u\colon R \to R$ where $\sigma_u(x) = uxu^{-1}$. The reader should verify that σ_u is a homomorphism of rings and is a one-to-one correspondence.
(3) Suppose R is a commutative ring, H and G are groups and $\theta\colon H \to G$ is a homomorphism of groups. The action $rh \mapsto r\theta(h)$ induces a homomorphism of group rings $R(H) \to R(G)$ (see Example 3.1.6).

(a) The homomorphism $\langle e\rangle \to G$ induces a homomorphism $\theta : R \to R(G)$. Notice that θ is one-to-one and the image of θ is contained in the center of $R(G)$.
(b) The homomorphism $G \to \langle e\rangle$ induces $\varepsilon : R(G) \to R$. Notice that η is onto, and the kernel of η contains the set of elements $D = \{1 - g \mid g \in G\}$. The reader should verify that the kernel of η is the ideal generated by D in $R(G)$ (see Definition 3.2.6). Sometimes ε is called the *augmentation map*.

(4) If R is a ring, then the zero mapping $R \to (0)$ is a homomorphism of rings. (In the category of rings, (0) is a terminal object.)
(5) If R is a ring, there is a unique homomorphism $\chi : \mathbb{Z} \to R$. In fact, by definition $\chi(1) = 1$ so $\chi(n) = n\chi(1) = n1$ for an arbitrary integer n. (In the category of rings, $\mathbb{Z}$ is an initial object.) The image of χ is the smallest subring of R. If R is a domain, the image of χ is called the *prime ring of* R. The kernel of χ is a subgroup of $\mathbb{Z}$, hence is equal to (n) for some nonnegative integer n. We call n the *characteristic* of R and write $n = \text{char}(R)$.

PROPOSITION 3.2.5. *Let $\phi : R \to S$ be a homomorphism of rings. Let J be an ideal in S. Then the following are true:*

(1) $\phi^{-1}(J)$ is an ideal in R.
(2) If ϕ is onto and A is an ideal of R, then $\phi(A)$ is an ideal of S.

PROOF. (1): We know from group theory that $\phi^{-1}(J), +$ is a subgroup of $R, +$ (see Exercise 2.3.15). Let $x \in \phi^{-1}(J)$, $r \in R$. Then $\phi(rx) = \phi(r)\phi(x) \in J$ since $\phi(x) \in J$. Therefore, $rx \in \phi^{-1}(J)$. Likewise, $xr \in \phi^{-1}(J)$.

(2): We know from group theory that $\phi(A), +$ is a subgroup of $S, +$ (see Exercise 2.3.15). Let $y \in \phi(A)$ and $s \in S = \phi(R)$. Then there exist $r \in R$ and $x \in A$ such that $s = \phi(r)$ and $y = \phi(x)$. Because $rx \in A$, we have $sy = \phi(r)\phi(x) = \phi(rx) \in \phi(A)$. Likewise, $ys \in \phi(A)$. So $\phi(A)$ is an ideal in S. □

DEFINITION 3.2.6. Let R be any ring and $X \subseteq R$. The *left ideal generated by* X is

$$\left\{ \sum_{i=1}^{n} r_i x_i \mid n \geq 1,\ r_i \in R,\ x_i \in X \right\}.$$

The reader should verify that the left ideal generated by X is equal to the intersection of the left ideals containing X. The *ideal generated by* X is

$$\left\{ \sum_{i=1}^{n} r_i x_i s_i \mid n \geq 1,\ r_i, s_i \in R,\ x_i \in X \right\}.$$

The reader should verify that the ideal generated by X is equal to the intersection of the ideals containing X. If A and B are left ideals of R, then $A + B$ is the set $\{a + b \mid a \in A,\ b \in B\}$. The left ideal generated by the set $\{ab \mid a \in A,\ b \in B\}$ is denoted AB. A left ideal (or ideal) is *principal* if it is generated by a single element. If I is generated by X, we write $I = (X)$. A commutative ring R is called a *principal ideal ring* if every ideal is a principal ideal. A *principal ideal domain* is an integral domain in which every ideal is principal. Sometimes we say R is a PID.

PROPOSITION 3.2.7. *Let R be any ring. If A and B are left ideals in R, then the following are true.*

(1) $A + B$ is a left ideal of R. If A and B are ideals, then $A + B$ is an ideal.

(2) *$A+B$ is the left ideal of R generated by the set $A \cup B$.*
(3) *$AB = \{\sum_{i=1}^{n} x_i y_i \mid n \geq 1,\ x_i \in A,\ y_i \in B\}$. If A and B are ideals, then AB is an ideal.*
(4) *If $X = \{a_1, \ldots, a_n\}$ is a finite subset of R, then (X), the ideal generated by X, is equal to $(a_1) + \cdots + (a_n)$.*

PROOF. The proof is left to the reader. □

EXAMPLE 3.2.8. Additional examples of ideals are listed here.

(1) In any ring, the set (0) is an ideal.
(2) In any ring R, if u is invertible, then for any $r \in R$ we see that $r = (ru^{-1})u$ is in the left ideal generated by u. That is, $(u) = R$. We call R the *unit ideal* of R. In R, the *trivial ideals* are (0) and R. If R is a division ring, the only left ideals in R are the trivial ideals.
(3) The ideals in $\mathbb{Z}$ are precisely the subgroups of $(\mathbb{Z}, +)$. That is, I is an ideal of $\mathbb{Z}$ if and only if $I = (n)$ for some n. The ring $\mathbb{Z}$ is a principal ideal domain.

EXAMPLE 3.2.9. Let k be a field and $R = k[w,x,y,z]$ the polynomial ring in four variables over k. Let $A = (w,x)$ and $B = (y,z)$. Then $wy + xz \in AB$, but $wy + xz$ cannot be factored as uv, where $u \in A$ and $v \in B$. This shows that in general the set $\{uv \mid u \in A,\ v \in B\}$ is not an ideal.

EXAMPLE 3.2.10. Let k be a field. The only ideals in k are the trivial ideals, by Example 3.2.8. In this example we prove that $R = M_2(k)$, the ring of 2-by-2 matrices over k has no proper two-sided ideal. The same proof can be modified to show $M_n(k)$ has no proper ideal for any $n \geq 1$ (see Exercise 3.2.33). Let $I \neq (0)$ be an ideal in R. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a nonzero element of I. After multiplying A by suitable permutation matrices if necessary, we can assume $a \neq 0$. Let e_{ij} denote the elementary matrix with 1 in row i column j, and 0 elsewhere. Then $e_{11}Ae_{11} = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix} \in I$. Multiplying by a^{-1} shows $e_{11} \in I$. Let $P_{12} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ be the permutation matrix. Then $P_{12}e_{11} = e_{21} \in I$, $e_{11}P_{12} = e_{12} \in I$, and $P_{12}e_{12} = e_{22} \in I$. This proves I contains $\{e_{11}, e_{12}, e_{21}, e_{22}\}$ which is a k-vector space basis for R. Hence, $I = R$.

EXAMPLE 3.2.11. Let F be a field and $M_2(F)$ the ring of 2-by-2 matrices over F. The reader should verify that $\left\{ \begin{pmatrix} a & 0 \\ c & 0 \end{pmatrix} \mid a, c \in F \right\}$ is the principal left ideal in $M_2(F)$ generated by the elementary matrix $e_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$. The principal right ideal generated by e_{21} is $\left\{ \begin{pmatrix} 0 & 0 \\ c & d \end{pmatrix} \mid c, d \in F \right\}$.

LEMMA 3.2.12. *Let R be any ring and $a \in R$. The following are equivalent.*

(1) a has a left inverse in R.
(2) $1 \in Ra$.
(3) $Ra = R$.

PROOF. (1) implies (2): We have $a^{-1} \in R$ such that $1 = a^{-1}a$.
(2) implies (3): We have $1 = ra$ for some $r \in R$. For each $x \in R$, $(xr)a = x(ra) = x \in Ra$.
(3) implies (1): $1 \in R = Ra$ implies $1 = ra$ for some $r \in R$. □

In Lemma 3.2.13 we list the fundamental properties of two-sided ideals in a ring R. It is the counterpart for ideals of Lemma 2.3.4. By R/I we denote the set of all left cosets of $I,+$ in $R,+$. Then the factor group R/I is an abelian group under addition and the natural map $\eta : R \to R/I$ is a homomorphism of additive groups.

LEMMA 3.2.13. *Let R be a ring and I a left ideal in R. The following are equivalent.*

(1) I is a two-sided ideal of R. That is, for each $r \in G$ and $x \in I$, we have $rx \in I$ and $xr \in I$.
(2) There is a well defined multiplicative binary operation $R/I \times R/I \to R/I$ on R/I defined by the rule $(x+I, y+I) \mapsto xy+I$.
(3) There is a multiplicative binary operation on R/I such that the natural map $\eta : R \to R/I$ is a homomorphism of rings.
(4) There exists a ring S and a homomorphism of rings $\theta : R \to S$ such that $I = \ker\theta$.

PROOF. (1) implies (2): We verify that multiplication of cosets is well defined. Say $x \equiv x' \pmod I$ and $y \equiv y' \pmod I$. Then $x - x' \in I$ implies that $xy - x'y = (x-x')y \in I$. Likewise $y - y' \in I$ implies that $x'y - x'y' = x'(y-y') \in I$. Taken together, we have $xy \equiv x'y \equiv x'y' \pmod I$.

(2) implies (3): On R/I, the associative law for multiplication, the distributive laws and the fact that $1+I$ is the multiplicative identity are routine to check. Therefore, R/I is a ring. Let $\eta : R \to R/I$ be the natural map defined by $x \mapsto x+I$. Then η is a homomorphism, $\operatorname{im}\eta = R/I$, and $\ker\eta = I$.

(3) implies (4): Take S to be R/I and for θ take the natural map η.

(4) implies (1): Let $x \in \ker\theta = I$ and $r \in R$. Then $\theta(rx) = \theta(r)\theta(x) = \theta(r)0 = 0$, by Proposition 3.1.2. Likewise, $\theta(xr) = \theta(x)\theta(r) = 0\theta(r) = 0$. This prove that xr and rx are in $\ker\theta = I$. □

DEFINITION 3.2.14. Let R be a ring and I an ideal in R. The *residue class ring* is the set $R/I = \{a+I \mid a \in R\}$ of all left cosets of I in R. We sometimes call R/I the factor ring, or quotient ring of R modulo I. We define addition and multiplication of cosets by the rules

$$(a+I)+(b+I) = (a+b)+I$$
$$(a+I)(b+I) = ab+I.$$

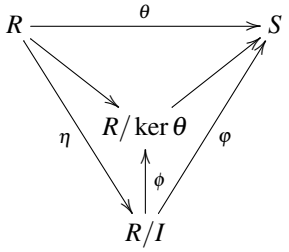
By Lemma 3.2.13, R/I is a ring, the natural map $\eta : R \to R/I$ is a homomorphism of rings, η is onto, and $I = \ker\eta$.

Theorem 3.2.15 and Corollaries 3.2.16 and 3.2.17 are the counterparts for rings of Theorems 2.3.11, 2.3.12 and 2.3.13.

THEOREM 3.2.15. *(The Fundamental Theorem on Ring Homomorphisms) Let $\theta : R \to S$ be a homomorphism of rings. Let I be an ideal of R contained in $\ker\theta$. There exists a homomorphism $\varphi : R/I \to S$ satisfying the following.*

(1) $\varphi(a+I) = \theta(a)$, or in other words $\theta = \varphi\eta$.
(2) φ is the unique homomorphism from $R/I \to S$ such that $\theta = \varphi\eta$.
(3) $\operatorname{im}\theta = \operatorname{im}\varphi$.
(4) $\ker\varphi = \eta(\ker\theta) = \ker(\theta)/I$.
(5) φ is one-to-one if and only if $I = \ker\theta$.
(6) φ is onto if and only if θ is onto.

(7) There is a unique homomorphism $\phi : R/I \to R/\ker\theta$ such that the diagram



commutes.

PROOF. On the additive groups, this follows straight from Theorem 2.3.11. The map φ is multiplicative since θ is a homomorphism of rings. □

COROLLARY 3.2.16. *Let R be a ring and $I \subseteq J \subseteq R$ a chain of ideals in R. Then J/I is an ideal in R/I and*

$$R/J \cong \frac{R/I}{J/I}.$$

PROOF. This follows from Theorem 3.2.15 and Theorem 2.3.12 (c). □

COROLLARY 3.2.17. *(Correspondence Theorem) Let R be a ring and I an ideal in R. There is a one-to-one order-preserving correspondence between the ideals J such that $I \subseteq J \subseteq R$ and the ideals of R/I given by $J \mapsto J/I$.*

PROOF. This follows from Theorem 3.2.15 and Theorem 2.3.13. □

2.1. Integral Domains. The next lemma and its proof are written using symbolic expressions.

LEMMA 3.2.18. *Let R be a ring in which $0 \neq 1$. The following are equivalent, where a,b,c represent elements of R.*

(1) $(ab = 0) \to ((a = 0) \vee (b = 0))$
(2) $(a \neq 0) \to (((ab = ac) \to (b = c)) \wedge ((ba = ca) \to (b = c)))$
(3) $((a \neq 0) \wedge (b \neq 0)) \to (ab \neq 0)$

PROOF. (1) is equivalent to (3) by contraposition.
(1) implies (2):

$$\begin{aligned}((a \neq 0) \wedge (ab = ac)) &\to ((a \neq 0) \wedge (ab - ac = 0)) \\ &\to ((a \neq 0) \wedge (a(b-c) = 0)) \\ &\to ((a \neq 0) \wedge ((a = 0) \vee (b = c))) \\ &\to (b = c)\end{aligned}$$

(2) implies (1): $((a \neq 0) \wedge (ab = 0)) \to (a(b-0) = 0) \to (ab = a0) \to b = 0.$ □

As in Definition 3.1.3, a ring that satisfies the three equivalent statements of Lemma 3.2.18 is a domain. A commutative domain is called an integral domain.

EXAMPLE 3.2.19. If F is a field, then F is an integral domain.

(1) If R is a subring of F, then R is an integral domain.

(2) The ring of 2-by-2 matrices $M_2(F)$ is a noncommutative F-algebra. Since $M_2(F)$ contains zero divisors, it is not a domain. For example:

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

THEOREM 3.2.20. *Let R be a finite integral domain. Then R is a field.*

PROOF. Let $a \in R - \{0\}$. Consider the "left multiplication by a" function $\ell_a : R \to R$. The distributive law on R implies $\ell_a(x+y) = a(x+y) = ax+ay = \ell_a(x)+\ell_a(y)$. Therefore, ℓ_a is a homomorphism on the additive group $(R,+)$. Since R is an integral domain, by ℓ_a is one-to-one, by Lemma 3.2.18. Since R is finite, the Pigeonhole Principle (Exercise 1.1.11) implies ℓ_a is onto. So there exists $x \in R$ such that $ax = 1$. This proves a is invertible. By Definition 3.1.3, R is a field. □

The proof of Theorem 3.2.20 shows that a finite domain is a division ring. By a theorem of Wedderburn ([**5**, Theorem 7.5.4]), a finite division ring is always commutative.

DEFINITION 3.2.21. Let R be a commutative ring. An ideal I in R is *prime* in case R/I is an integral domain. An ideal I in R is *maximal* in case R/I is a field. A field is an integral domain, so a maximal ideal is a prime ideal. By Definition 3.1.3, an integral domain has at least two elements, so the unit ideal is never prime.

EXAMPLE 3.2.22. In an integral domain, the zero ideal (0) is a prime ideal. In a commutative ring R, the zero ideal (0) is a maximal ideal if and only if R is a field (Exercise 3.2.32). Let P be a nonzero prime ideal in $\mathbb{Z}$. Then $\mathbb{Z}/P$ is a finite integral domain which is a field, by Theorem 3.2.20. The maximal ideals in $\mathbb{Z}$ are the nonzero prime ideals.

PROPOSITION 3.2.23. *Let R be a commutative ring and P an ideal of R. Assume $P \neq R$. The following are equivalent.*

(1) P is a prime ideal. That is, R/P is an integral domain.
(2) For all $x, y \in R$, if $xy \in P$, then $x \in P$ or $y \in P$.
(3) For any ideals I, J in R, if $IJ \subseteq P$, then $I \subseteq P$ or $J \subseteq P$.

PROOF. Is left to the reader. □

PROPOSITION 3.2.24. *Let $\phi : R \to S$ be a homomorphism of commutative rings. Let J be an ideal in S. Then the following are true:*

(1) If J is a prime ideal, then $\phi^{-1}(J)$ is a prime ideal.
(2) If ϕ is onto, and J is a maximal ideal, then $\phi^{-1}(J)$ is a maximal ideal.

PROOF. (1): This is Exercise 3.2.47.

(2): Let J be a maximal ideal of S. By (2) we know $\phi^{-1}(J) \neq R$. Assume A is an ideal of R such that $\phi^{-1}(J) \subseteq A \subseteq R$. Because ϕ is onto, we have $J = \phi\phi^{-1}(J) \subseteq \phi(A) \subseteq \phi(R) = S$. By (4), $\phi(A)$ is an ideal of S. Since J is maximal, $\phi(A) = J$ or $\phi(A) = S$. First suppose $\phi(A) = J$. Then $A \subseteq \phi^{-1}\phi(A) = \phi^{-1}(J)$. So $\phi^{-1}(J) = A$. Now suppose $\phi(A) = S$. Then $1 \in \phi(A)$. Say $u \in A$ and $\phi(u) = 1$. Since $\phi(1) = 1$, we have $\phi(1-u) = 0$. So $1-u \in \ker\phi$. But $\ker\phi = \phi^{-1}(0) \subseteq \phi^{-1}(J) \subseteq A$. So $1 = u-(u-1) \in A$, which implies $A = R$. This proves $\phi^{-1}(J)$ is a maximal ideal. □

COROLLARY 3.2.25. *(Correspondence Theorem for Prime Ideals) Let R be a commutative ring and I an ideal in R. There is a one-to-one order-preserving correspondence between the ideals J such that $I \subseteq J \subseteq R$ and the ideals of R/I given by $J \mapsto J/I$. Under this correspondence prime ideals of R/I correspond to prime ideals of R that contain I.*

PROOF. The first part is Corollary 3.2.17. The preimage of a prime ideal is a prime ideal, by Proposition 3.2.24 (1). Corollary 3.2.16 shows that the image of a prime ideal that contains I is a prime ideal in R/I. □

PROPOSITION 3.2.26. *Let R be a commutative ring.*

(1) An ideal M is a maximal ideal in R if and only if M is not contained in a larger proper ideal of R.
(2) R contains a maximal ideal.
(3) If I is a proper ideal of R, then R contains a maximal ideal M such that $I \subseteq M$.

PROOF. (1): By Exercise 3.2.32 and Corollary 3.2.17 R/M is a field if and only if there is no proper ideal J such that $M \subsetneq J$.

(2): Let $\mathscr{S}$ be the set of all ideals I in R such that $I \neq R$. Then $(0) \in \mathscr{S}$. Order $\mathscr{S}$ by set inclusion. Let $\{A_\alpha\}$ be a chain in $\mathscr{S}$. The union $J = \bigcup A_\alpha$ is an ideal in R, by Exercise 3.2.34. Since 1 is not in any element of $\mathscr{S}$, it is clear that $1 \notin J$. Therefore, $J \in \mathscr{S}$ is an upper bound for the chain $\{A_\alpha\}$. By Zorn's Lemma, Proposition 1.3.3, $\mathscr{S}$ contains a maximal member. By Part (1), this ideal is a maximal ideal. □

2.2. Exercises.

EXERCISE 3.2.27. Prove that if $\theta : R \to S$ is a homomorphism of rings, then the image of θ is a subring of S and the kernel of θ is a two-sided ideal of R.

EXERCISE 3.2.28. Let $\theta : R \to S$ be a homomorphism of rings. Prove:

(1) θ is one-to-one if and only if $\ker\theta = (0)$.
(2) If R is a division ring, then θ is one-to-one.

EXERCISE 3.2.29. Let R be any ring.

(1) If $n = \operatorname{char} R$, then $nx = 0$ for any $x \in R$.
(2) If R is a domain, then the characteristic of R is either 0 or a prime number.

EXERCISE 3.2.30. Let R be any ring and suppose $p = \operatorname{char} R$ is a prime number. Let x and y be elements of R such that $xy = yx$. Prove:

(1) $(x+y)^p = x^p + y^p$.
(2) $(x-y)^p = x^p - y^p$.
(3) $(x-y)^{p-1} = \sum_{i=0}^{p-1} x^i y^{p-1-i}$.
(4) If $n \geq 0$, then $(x+y)^{p^n} = x^{p^n} + y^{p^n}$.

(Hint: Exercise 1.2.21.) See Exercise 3.6.31 for an application of this exercise.

EXERCISE 3.2.31. Let R be a commutative ring and assume $\operatorname{char} R = p$ is a prime number. Define $\theta : R \to R$ by $x \mapsto x^p$. Show that θ is a homomorphism of rings. We call θ the *Frobenius homomorphism*. For any $a \geq 1$, show that $\theta^a(x) = x^{p^a}$. If R is a field, show that θ is one-to-one.

EXERCISE 3.2.32. Prove:

(1) If R is a ring with no proper left ideal, then every nonzero element has a left inverse. (Hint: Exercise 2.1.22.)
(2) If R is a ring with no proper left ideal, then R is a division ring. (Hint: $R - (0)$ is a monoid.)
(3) A commutative ring R is a field if and only if R has no proper ideal.

EXERCISE 3.2.33. This exercise is a continuation of Example 3.2.10. Let R be a ring and $M_n(R)$ the ring of n-by-n matrices over R where addition and multiplication are defined in the usual way.

(1) Let e_{ij} be the elementary matrix which has 0 in every position except in position (i, j) where there is 1. Determine the left ideal in $M_n(R)$ generated by e_{ij}.
(2) If $n \geq 2$, show that $M_n(R)$ has proper left ideals.
(3) If I is an ideal in $M_n(R)$, show that $I = M_n(J)$ for some ideal J in R. (Hint: Use multiplication by the various E_{ij}.)
(4) If D is a division ring, show that $M_n(D)$ has no proper ideal. We say that $M_n(D)$ is a *simple ring*.

EXERCISE 3.2.34. Let R be a ring, I an index set, and $\{A_i \mid i \in I\}$ a family of left ideals in R.

(1) Show that $\bigcap_{i\in I} A_i$ is a left ideal in R.
(2) Suppose $\{A_i \mid i \in I\}$ is an ascending chain of left ideals in R. That is, I is a partially ordered set that is a chain, and if $\alpha \leq \beta$ in I, then $A_\alpha \subseteq A_\beta$. Show that $\bigcup_{i\in I} A_i$ is a left ideal in R.

EXERCISE 3.2.35. Let U and V be ideals in the commutative ring R. As in Definition 3.2.6, UV is the ideal generated by the set $\{uv \mid u \in U, v \in V\}$. Prove the following.

(1) $UV \subseteq U \cap V$.
(2) If $U + V = R$, then $UV = U \cap V$.
(3) Show by counterexample that $UV = U \cap V$ is false in general.

EXERCISE 3.2.36. Let $n > 1$.

(1) Show that every prime ideal in $\mathbb{Z}/(n)$ is a maximal ideal.
(2) Let $n = \pi_1^{e_1} \cdots \pi_k^{e_k}$ be the unique factorization of n (Proposition 1.2.7). Determine the maximal ideals in $\mathbb{Z}/(n)$.

EXERCISE 3.2.37. An element x of a ring is said to be *nilpotent* if $x^n = 0$ for some $n > 0$. If R is a commutative ring, let $\mathrm{Rad}_R(0)$ denote the set of all nilpotent elements of R. We call $\mathrm{Rad}_R(0)$ the *nil radical* of R.

(1) Show that $\mathrm{Rad}_R(0)$ is an ideal.
(2) Let I be an ideal of R contained in $\mathrm{Rad}_R(0)$. Show that the nil radical of R/I is $\mathrm{Rad}_R(0)/I$, hence the nil radical of $R/\mathrm{Rad}_R(0)$ is the trivial ideal $(0 + \mathrm{Rad}_R(0))$.

EXERCISE 3.2.38. Let $\theta : R \to S$ be a homomorphism of rings. Prove that θ induces a homomorphism $\theta : \mathrm{Units}(R) \to \mathrm{Units}(S)$ on the groups of units.

EXERCISE 3.2.39. Let R be a commutative ring, $\mathrm{Rad}_R(0)$ the nil radical of R, and $\eta : R \to R/\mathrm{Rad}_R(0)$ the natural map. Prove:

(1) If x is a nilpotent element of R, then $1 + x$ is a unit in R.
(2) If $\eta(r)$ is a unit in $R/\mathrm{Rad}_R(0)$, then r is a unit in R.
(3) If I is an ideal of R contained in $\mathrm{Rad}_R(0)$, then the natural map $\eta : \mathrm{Units}(R) \to \mathrm{Units}(R/I)$ is onto and the kernel of η is equal to the coset $1 + I$.

EXERCISE 3.2.40. Let I and J be ideals in the commutative ring R. The *ideal quotient* is $I : J = \{x \in R \mid xJ \subseteq I\}$. Prove that $I : J$ is an ideal in R.

EXERCISE 3.2.41. For the following, let I, J and K be ideals in the commutative ring R. Prove that the ideal quotient satisfies the following properties.

(1) $I \subseteq I : J$

(2) $(I:J)J \subseteq I$
(3) $(I:J):K = I:JK = (I:K):J$
(4) If $\{I_\alpha \mid \alpha \in S\}$ is a collection of ideals in R, then

$$\left(\bigcap_{\alpha\in S} I_\alpha\right) : J = \bigcap_{\alpha\in S} (I_\alpha : J)$$

(5) If $\{J_\alpha \mid \alpha \in S\}$ is a collection of ideals in R, then

$$I : \sum_{\alpha\in S} J_\alpha = \bigcap_{\alpha\in S} (I : J_\alpha)$$

EXERCISE 3.2.42. A *local ring* is a commutative ring R such that R has exactly one maximal ideal. If R is a local ring with maximal ideal $\mathfrak{m}$, then $R/\mathfrak{m}$ is called the *residue field* of R. If $(R,\mathfrak{m})$ and $(S,\mathfrak{n})$ are local rings and $f : R \to S$ is a homomorphism of rings, then we say f is a *local homomorphism of local rings* in case $f(\mathfrak{m}) \subseteq \mathfrak{n}$. Prove:

(1) A field is a local ring.
(2) If $(R,\mathfrak{m})$ is a local ring, then the group of units of R is equal to the set $R - \mathfrak{m}$.
(3) If $f : R \to S$ is a local homomorphism of local rings, then f induces a homomorphism of residue fields $R/\mathfrak{m} \to S/\mathfrak{n}$.

EXERCISE 3.2.43. Let R be a ring. If A and B are left ideals in R, then the product ideal AB is defined in Definition 3.2.6. The powers of A are defined recursively by the rule:

$$A^n = \begin{cases} R & \text{if } n = 0, \\ A & \text{if } n = 1, \\ AA^{n-1} & \text{if } n > 1. \end{cases}$$

The left ideal A is *nilpotent* if for some $n > 0$, $A^n = 0$. Let A and B be nilpotent left ideals of R. Prove:

(1) Assume $A^n = 0$. If $x_1,\ldots,x_n$ are elements of A, then $x_1 \cdots x_n = 0$.
(2) Every element x of A is nilpotent.
(3) $A + B$ is a nilpotent left ideal. (Hint: For all p sufficiently large, if $x_1,\ldots,x_p$ are elements of $A \cup B$, show that $x_1 \cdots x_p = 0$.)

EXERCISE 3.2.44. Let R be a commutative ring and $\{x_1,\ldots,x_n\}$ a finite set of nilpotent elements of R. Show that $Rx_1 + \cdots + Rx_n$ is a nilpotent ideal.

EXERCISE 3.2.45. Let R be a ring. We say that a left ideal M of R is *maximal* if M is not equal to R and if I is a left ideal such that $M \subseteq I \subsetneq R$, then $M = I$. Let I be a left ideal of R which is not the unit ideal. Apply Zorn's Lemma, Proposition 1.3.3, to show that there exists a maximal left ideal M such that $I \subseteq M \subsetneq R$.

EXERCISE 3.2.46. Prove Proposition 3.2.23.

EXERCISE 3.2.47. Prove Proposition 3.2.24 (1).

EXERCISE 3.2.48. If R is a commutative ring, let $\mathrm{Aut}(R)$ denote the group of all ring automorphisms of R. Prove the following.

(1) $\mathrm{Aut}(\mathbb{Z}) = (1)$.
(2) $\mathrm{Aut}(\mathbb{Z}/(n)) = (1)$ for any n.

EXERCISE 3.2.49. Let R be a commutative ring and G a group. Show that the group ring $R(G)$ (see Example 3.1.6) is isomorphic to the opposite ring $R(G)^o$ (see Definition 3.1.8). (Hints: Exercise 2.1.24 and Example 3.2.4 (3).)

3. Direct Product and Direct Sum of Rings

In the definitions and theorems of this section the direct product and direct sum of rings are limited to two factors and two summands. This restriction is for the sake of simplicity. All of the following results can be generalized to products and sums involving an arbitrary finite number of terms.

DEFINITION 3.3.1. Let R and S be rings. The *direct product of R and S* is the ring with underlying set $R \times S$ where addition and multiplication are defined coordinate-wise:

$$(a,b)+(c,d) = (a+c, b+d)$$
$$(a,b)(c,d) = (ac, bd).$$

Since R and S each contain an additive identity denoted 0, the additive identity in the product $R \times S$ is the ordered pair $(0,0)$. Since R and S both contain a multiplicative identity denoted 1, the multiplicative identity in the product is the ordered pair $(1,1)$. The reader should verify that the associative laws and distributive laws are satisfied.

DEFINITION 3.3.2. Let R be a ring. An *idempotent* of R is an element $e \in R$ that satisfies the equation $e^2 = e$. The elements 0 and 1 are called the trivial idempotents.

LEMMA 3.3.3. *Let R and S be rings.*

(1) Let $e_1 = (1,0)$ and $e_2 = (0,1)$ in $R \times S$. Then $e_1^2 = e_1$, $e_2^2 = e_2$, $e_1 e_2 = 0$, $e_1 \in Z(R \times S)$, $e_2 \in Z(R \times S)$, and $(1,1) = e_1 + e_2$. We say $\{e_1, e_2\}$ is a set of orthogonal idempotents for $R \times S$.

(2) The canonical projection maps define homomorphisms of rings: $\pi_1 : R \times S \to R$, where $\pi_1(a,b) = a$, and $\pi_2 : R \times S \to S$, where $\pi_2(a,b) = b$. Both π_1 and π_2 are onto. The kernel of π_1 is $(0) \times S$, which is the principal ideal generated by e_2. The kernel of π_2 is $R \times (0)$, which is the principal ideal generated by e_1.

(3) The canonical injection maps are $\iota_1 : R \to R \times S$, where $\iota_1(a) = (a,0)$, $\iota_2 : S \to R \times S$, where $\iota_2(b) = (0,b)$. Then each ι_i is a one-to-one homomorphism of additive groups. Moreover, ι_i is multiplicative and $\pi_1 \iota_1 = 1_R$, $\pi_2 \iota_2 = 1_S$.

PROOF. The proof is left as an exercise for the reader. □

DEFINITION 3.3.4. Let I and J be proper ideals in a ring R. We say that R is the *internal direct sum* of I and J in case

(1) $R = I + J$, and
(2) for each $x \in R$, x has a unique representation as a sum $x = a + b$ where $a \in I$ and $b \in J$.

We denote the internal direct sum by $R = I \oplus J$. Notice that in this case the additive group $R,+$ is the internal direct product of the subgroups $I,+$ and $J,+$, but it is customary to say direct sum instead of direct product when the group is written additively.

DEFINITION 3.3.5. If R is a ring and I and J are ideals in R, then we say I and J are *comaximal* if $I + J = R$.

THEOREM 3.3.6. *(Fundamental Theorem on Internal Direct Sums of Ideals) If I and J are ideals in the ring R and $R = I \oplus J$, then the following are true.*

(1) $I \cap J = (0)$.

(2) If $x \in I$ and $y \in J$, then $xy = yx = 0$.

(3) I is a ring and J is a ring. Let e_1 denote the identity element of I and e_2 the identity for J. Then $\{e_1, e_2\}$ is a set of orthogonal idempotents in R. Each e_i is in the center of R. $I = Re_1$ and $J = Re_2$ are principal ideals in R.

(4) R is isomorphic to the (external) direct product $I \times J$.
(5) Suppose L is a left ideal in the ring I and K is a left ideal in J. Then $L+K$ is a left ideal in R, and the sum $L+K$ is a direct sum.
(6) If U is a left ideal of R, then $U = L \oplus K$ where L is a left ideal in the ring I and K is a left ideal in the ring J.

PROOF. (1): Since $R,+$ is an additive group with subgroups $I,+$ and $J,+$, Part (1) follows from the corresponding statement about an internal direct product of a group.

(2): Notice that xy and yx are both in $I \cap J$ since the ideals are two-sided.

(3): Because I is an ideal, it is enough to show that I has a multiplicative identity. Write $1 = e_1 + e_2$. If $x \in I$, then multiply by x from the left and use Part (2) to get $x = x1 = xe_1 + xe_2 = xe_1$. Now multiply by x from the right and use Part (2) to get $x = 1x = e_1x + e_2x = e_1x$. This shows e_1 is the multiplicative identity for I. Likewise, e_2 is the multiplicative identity for J. Orthogonality of $\{e_1, e_2\}$ is by Part (2). The rest is left to the reader.

(4): Define a function $f : I \times J \to R$ from the external ring direct product to R by the rule $(x,y) \mapsto x+y$. By the corresponding statement about an internal direct product of a group, f is an isomorphism on additive groups. The reader should verify using Part (2) that f is multiplicative.

(5): Since each element r in $R = I + J$ has a unique representation in the form $r = r_1 + r_2$, so does any element x in $I = L + K$. So the sum is a direct sum and we can write $x = x_1 + x_2$ where $x_1 \in L$ and $x_2 \in K$ are unique. Then $rx = r_1x_1 + r_2x_2$ is in $L+K$, which shows $L+K$ is a left ideal in R.

(6): By Part (3), there are central idempotents e_1 and e_2 in R such that $I = Re_1$ and $J = Re_2$. Let $L = e_1U$ and $K = e_2U$. Since e_1 and e_2 are central, $L = Ue_1$ and $K = Ue_2$ are left ideals in R. Since $U \subseteq R$ we have $L = Ue_1 \subseteq Re_1 = I$, so L is a left ideal in I. Likewise, $K = Ue_2 \subseteq Re_2 = J$, so K is a left ideal in J. Since $1 = e_1 + e_2$, we see that $U = L + K$. The sum is a direct sum by Part (5). □

THEOREM 3.3.7. *(The Chinese Remainder Theorem) Let R be a ring and I, J comaximal ideals of R. Then*

$$\frac{R}{I \cap J} \cong \frac{R}{I} \times \frac{R}{J}$$

where the isomorphism is induced by the natural projections $\eta_1 : R \to R/I$ and $\eta_2 : R \to R/J$.

PROOF. Step 1: Let $\phi : R \to R/I \times R/J$ be defined by $\phi(x) = (x+I, x+J)$. Since ϕ is defined in terms of the natural projections η_1, η_2, ϕ is a well defined homomorphism of rings.

Step 2: We prove that ϕ is onto. Let $a,b \in R$. We need to find $x \in R$ such that $\phi(x) = (a+I, b+J)$. Since I and J are comaximal, there exist $u \in I$, $v \in J$ such that $1 = u + v$. Then $u = 1 - v \equiv 1 \pmod{J}$ and $v = 1 - u \equiv 1 \pmod{I}$. Set $x = bu + av$. Then

$$\begin{aligned} x &\equiv bu + av \pmod{I} \\ &\equiv av \pmod{I} \\ &\equiv a \pmod{I}. \end{aligned}$$

Likewise, $x \equiv b \pmod{J}$. Therefore, $\phi(x) = (a+I, b+J)$.

Step 3: Consider the kernel of ϕ, $\ker\phi = \{x \in R \mid x \in I \text{ and } x \in J\} = I \cap J$. By Theorem 3.2.15, this proves the theorem. □

PROPOSITION 3.3.8. *Let R be a commutative ring. If I and J are comaximal ideals, then $IJ = I \cap J$.*

PROOF. If $x \in I$ and $y \in J$, then $xy \in I$ and $xy \in J$. Since IJ is generated by elements of the form xy, we have $IJ \subseteq I \cap J$. Let z be an arbitrary element of $I \cap J$. We show $z \in IJ$. Since $R = I + J$, there exist $u \in I$ and $v \in J$ such that $1 = u + v$. Now $zu \in IJ$ since $z \in J$ and $u \in I$. Also $zv \in IJ$ since $z \in I$ and $v \in J$. Then $z = zu + zv \in IJ$. □

COROLLARY 3.3.9. *Let R be a commutative ring. If I and J are comaximal ideals, then $R/IJ \cong R/I \times R/J$.*

EXAMPLE 3.3.10. Let F be a field and

$$R = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \mid a,b,d \in F \right\}$$

the set of all upper triangular matrices in $M_2(F)$. As in Example 3.1.12, R is a noncommutative subring of $M_2(F)$. The proof given in Example 3.1.13 can be used to show that the center of R is the set of scalar matrices, which is isomorphic to F by the homomorphism $\delta : F \to R$ defined by $\delta(a) = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$. Define $\lambda : R \to F$ by $\lambda \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = a$. The reader should verify that λ is a homomorphism and $\lambda\delta(a) = a$ for all $a \in F$. We say F is a subfield of R and λ is a *section to* δ. The homomorphism $\rho : R \to F$ defined by $\rho \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = d$ also satisfies $\rho\delta(a) = a$, hence a section to δ is not unique. The kernels of λ and ρ are

$$\ker\lambda = \left\{ \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} \mid b,d \in F \right\}, \quad \ker\rho = \left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \mid a,b \in F \right\},$$

which are proper ideals in R. We say R is not a simple ring. Since F has no proper ideals, by Corollary 3.2.17, there is no proper ideal of R that contains $\ker\lambda$ or $\ker\rho$. The ideals $\ker\lambda$ and $\ker\rho$ are maximal proper ideals in R. Let $D = \left\{ \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \mid a,d \in F \right\}$. The reader should verify that D is a subring of R. Define $\tau : R \to D$ by $\tau \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$. The reader should verify that τ is a homomorphism and for any matrix $A \in D$, $\tau(A) = A$. In other words, τ is a section to the inclusion map $D \to R$. The kernel of τ is the ideal

$$\ker\tau = \left\{ \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \mid b \in F \right\}.$$

If $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ is an idempotent matrix, then a and d are idempotents in F. After looking at the possible cases, the reader should verify that the set of all idempotents in R is

$$\left\{ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\}.$$

Only the two trivial idempotents, namely 0 and 1, are central. Therefore, R is not an internal direct sum of proper ideals. Let R^* be the group of units of R. By Exercise 3.2.38, there are homomorphisms of groups $\delta^* : F^* \to R^*$ and $\rho^* : R^* \to F^*$. Let

$$T = \ker\rho^* = \left\{ \begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix} \mid a \in F^*, b \in F \right\},$$

and

$$Z = \delta(F^*) = \left\{ \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \mid a \in F^* \right\}.$$

By Exercise 2.5.21, the group of units of R is the internal direct product $R^* = T \times Z$ of the two proper normal subgroups T and Z. The ring R is an example of an *extension of a ring by a module*. Specifically, R is the extension of D by the module $\ker \tau$. The interested reader is referred to [**5**, Exercise 8.1.14] for the general construction.

3.1. Exercises.

EXERCISE 3.3.11. Suppose R is a ring and $e \in R$ is a central idempotent. Assume $e \neq 0$ and $e \neq 1$. Let I be the ideal generated by e. Prove that R is equal to the internal direct sum $I \oplus J$ for some ideal J.

EXERCISE 3.3.12. Let k be a field of characteristic different from 2. Let $f = x^2 - 1$. Show that $k[x]/(f)$ is isomorphic to a direct sum of fields.

EXERCISE 3.3.13. Consider the ring $R = \mathbb{Z}/(n)$.

(1) Suppose $n = 1105$.
   (a) Prove that R is isomorphic to a direct sum of fields.
   (b) Determine all maximal ideals in R.
   (c) Determine all idempotents in R.
(2) Suppose $n = 1800$.
   (a) Determine all maximal ideals in R.
   (b) Determine all idempotents in R.

EXERCISE 3.3.14. Assume the ring R is the direct sum $R = R_1 \oplus R_2$. Let e_1, e_2 be the central idempotents corresponding to the direct summands (guaranteed by Theorem 3.3.6 (3)). Let D be a ring which has exactly two idempotents, namely 0 and 1. Let $\theta : R \to D$ be a homomorphism of rings. Prove that exactly one of the following is true:

(1) $\theta(e_1) = 1$ and $\theta(e_2) = 0$, or
(2) $\theta(e_1) = 0$ and $\theta(e_2) = 1$.

EXERCISE 3.3.15. Let R be any ring. Let I and J be ideals in R and $\phi : R \to R/I \oplus R/J$ the natural homomorphism of Theorem 3.3.7. Show that the image of ϕ is the subring of $R/I \oplus R/J$ defined by $\{(x+I, y+J) \mid x - y \in I + J\}$. See [**4**, Exercise 4.1.45] for an interpretation of this result in terms of modules.

EXERCISE 3.3.16. If $n > 1$, then we say n is square free if n is not divisible by the square of a prime number. Prove that the nil radical of $\mathbb{Z}/n$ is (0) if and only if n is square free. For the definition of nil radical, see Exercise 3.2.37.

EXERCISE 3.3.17. Let $n > 1$ and R a finite ring of order n. Suppose n is square free and the factorization of n into primes is $n = p_1 \cdots p_m$. Prove the following:

(1) $R \cong \mathbb{Z}/n$.
(2) R is commutative.
(3) R is a field, or a direct sum of fields.
(4) In terms of the prime factors of n, describe the maximal ideals of R.

4. Factorization in Commutative Rings

DEFINITION 3.4.1. Let R be a commutative ring. Suppose a and b are elements of R. We say *a divides b*, and write $a \mid b$, in case there exists $c \in R$ such that $b = ac$. We also say that a is a *factor* of b, or b is a *multiple* of a.

DEFINITION 3.4.2. Let R be a commutative ring and suppose a and b are elements of R. If $a \mid b$ and $b \mid a$, then we say a and b are *associates*. In this case we write $a \sim b$. The reader should verify that the relation "a is an associate of b" is an equivalence relation on R.

LEMMA 3.4.3. *Let R be a commutative ring. Let $a, b, r, u \in R$.*

(1) The following are equivalent:
  (a) $a \mid b$.
  (b) $b \in Ra = (a)$.
  (c) $(a) \supseteq (b)$.
(2) a and b are associates if and only if $(a) = (b)$.
(3) If $a = bu$ and u is a unit, then a and b are associates.
(4) If R is an integral domain and a and b are associates, then $a = bu$ for some unit u.
(5) Let R be an integral domain. If $a \neq 0$ and $a \mid b$, then there exists a unique c such that $b = ac$. We write $c = ba^{-1}$, or $c = b/a$.

PROOF. (1): This follows straight from Definitions 3.2.6 and 3.4.1.

(5): Suppose $b = ac = ac'$. Subtract and distribute to get $a(c - c') = 0$. Since $a \neq 0$ and R is an integral domain, this means $c - c' = 0$, hence $c = c'$.

The rest of the proof is left to the reader. □

DEFINITION 3.4.4. Let R be a commutative ring and a an element of R which is not a unit and not a zero divisor. Then a is *irreducible* in case whenever $a = bc$, then either b is a unit or c is a unit. We say that a is *prime* in case whenever $a \mid bc$, then either $a \mid b$ or $a \mid c$.

LEMMA 3.4.5. *Let R be an integral domain.*

(1) $p \in R$ is prime if and only if (p) is a prime ideal.
(2) If p is prime, then p is irreducible.
(3) If p is irreducible and q is an associate of p, then q is irreducible.
(4) If p is prime and q is an associate of p, then q is prime.
(5) If p is irreducible, then the only divisors of p are units and associates of p.

PROOF. In the following, let $a, b, p, q, u \in R$.

(1): We have $ab \in (p)$ if and only if $p \mid ab$. Likewise, $a \in (p)$ if and only if $p \mid a$, and $b \in (p)$ if and only if $p \mid b$.

(2): Suppose p is prime and $p = ab$. Since p is prime we assume $p \mid a$. Therefore a and p are associates. By Lemma 3.4.3 (4), b is a unit in R.

(3): Is a homework exercise.

(4): Assume p is prime, u is a unit, $q = pu$, and $q \mid ab$. For some $c \in R$, $ab = qc = puc$. Since p is prime we assume $p \mid a$. For some $d \in R$, $a = pd = (pu)(u^{-1}d)$, which shows $q \mid a$.

(5): The proof is left to the reader. □

4.1. Greatest Common Divisors.

DEFINITION 3.4.6. Let R be a commutative ring and X a nonempty subset of R. An element $d \in R$ is a *greatest common divisor of X* if the following are satisfied:

(1) $d \mid x$ for all $x \in X$, and
(2) if $c \mid x$ for all $x \in X$, then $c \mid d$.

We sometimes write $d = \gcd(X)$ if d is a greatest common divisor of X. When $X = \{x_1, \ldots, x_n\}$ is finite, we write $d = \gcd(x_1, \ldots, x_n)$ for $\gcd(X)$. Notice that if d is a greatest common divisor, so is any associate of d. If $\gcd(X)$ exists, it is not unique.

LEMMA 3.4.7. *Let X be a nonempty subset of an integral domain R. If $d = \gcd(X)$ exists, then d is unique up to associates. That is, if d and d' are two greatest common divisors of X, then there exists a unit $u \in R^*$ such that $d' = du$, hence d and d' are associates.*

PROOF. By Definition 3.4.6, we have $d \mid d'$ and $d' \mid d$. Thus d and d' are associates. By Lemma 3.4.3 (4), $d' = du$ for some $u \in R^*$. □

PROPOSITION 3.4.8. *Let R be a commutative ring and X a nonempty subset of R.*

(1) If the ideal generated by X is principal and d is a generator for (X), then $d = \gcd(X)$.
(2) If $d = \gcd(X)$ exists and d is in the ideal (X), then $(d) = (X)$.

PROOF. (1): If $(d) = (X)$, then $d \mid x$, for all $x \in X$. Also, $d = a_1x_1 + \cdots + a_nx_n$ for some $a_1, \ldots, a_n \in R$ and $x_1, \ldots, x_n \in X$. Suppose $c \mid x$ for each $x \in X$. Then $c \mid a_1x_1 + \cdots + a_nx_n = d$.

(2): This follows from Definition 3.4.6 and Exercise 3.4.28. □

COROLLARY 3.4.9. *(A PID is a Bézout domain) If R is a PID, and X is a nonempty subset of R, then $d = \gcd(X)$, the greatest common divisor of X, exists and is unique up to associates. Any generator d of the ideal (X) is a greatest common divisor of a and b. In this case, $d = a_1x_1 + \cdots + a_nx_n$ for some $a_1, \ldots, a_n \in R$ and $x_1, \ldots, x_n \in X$.*

PROOF. Since (X) is principal, there exists $d \in R$ such that $(d) = (X)$. Proposition 3.4.8 (1) implies $d = \gcd(X)$ exists and can be written in the form $d = a_1x_1 + \cdots + a_nx_n$ for some $a_1, \ldots, a_n \in R$ and $x_1, \ldots, x_n \in X$. By Lemma 3.4.7, d is unique up to associates. □

COROLLARY 3.4.10. *Let R be a PID and $p \in R$ an irreducible element. Then the following are true.*

(1) p is prime. That is, if $p \mid ab$, then $p \mid a$ or $p \mid b$.
(2) If $x_1, x_2, \ldots, x_n$ in R and $p \mid x_1x_2 \cdots x_n$, then $p \mid x_i$ for some i.

PROOF. (1): Assume $p \mid ab$ and p does not divide b. We prove $p \mid a$. The ideal (p, b) is principal, hence is equal to (d), for some $d \in R$. Then $d \mid p$ and $d \mid b$. Since p is irreducible, d is a unit, or d is an associate of p (Lemma 3.4.5 (5)). We are assuming p does not divide b, hence d is not an associate of p, hence d is a unit. Therefore $(d) = (1)$. By Corollary 3.4.9, we can write $1 = px + by$. Multiply by a to get $a = pax + aby$. Since $p \mid ab$, this shows $p \mid a$.

(2) If $n = 1$, then take $i = 1$ and stop. Assume inductively that $n > 1$ and the result holds for a product of $n-1$ factors. Then $p \mid (x_1 \cdots x_{n-1})x_n$. By Part (1), $p \mid x_n$, or $p \mid (x_1 \cdots x_{n-1})$. By the induction hypothesis, $p \mid x_i$ for some i. □

DEFINITION 3.4.11. Let R be an integral domain. Then R is a *unique factorization domain* if for every nonzero nonunit x in R, the following are satisfied:

(1) x has a representation as a product of irreducibles. That is, there exist irreducible elements $x_1, x_2, \ldots, x_n$ in R such that $x = x_1x_2 \cdots x_n$.
(2) In any factorization of x as in (1), the number of factors is unique.
(3) In any factorization of x as in (1), the irreducible factors are unique up to order and associates.

Sometimes we say R is a UFD.

EXAMPLE 3.4.12. The ring $\mathbb{Z}$ is a UFD, by the Fundamental Theorem of Arithmetic. We will prove in Theorem 3.4.26 that any PID is a UFD.

COROLLARY 3.4.13. *Let R be a UFD. If $X = \{r_1, \ldots, r_n\}$ is a finite nonempty subset of R, then $d = \gcd(X)$ exists and is unique up to associates.*

PROOF. If $n = 1$, then by Proposition 3.4.8 (1), $r_1 = \gcd(X)$ exists. By Mathematical Induction and Exercise 3.4.29, it suffices to prove the $n = 2$ case. Assume $X = \{a, b\}$. If $a = 0$, then $(a, b) = (b)$ and by Proposition 3.4.8 (1), $b = \gcd(a, b)$ exists. If $(a, b) = (1)$, then by Proposition 3.4.8 (1), $1 = \gcd(a, b)$ exists. Assume a and b are both nonzero and nonunits. Then by Exercise 3.4.30, $\gcd(a, b)$ exists and we are done. □

COROLLARY 3.4.14. *Let R be a UFD and $p \in R - (0)$. Then the following are equivalent.*

(1) p is irreducible.
(2) p is prime.
(3) The principal ideal (p) is a prime ideal.

PROOF. By Lemma 3.4.5 (1), (2) is equivalent to (3). By Lemma 3.4.5 (2), (2) implies (1). We prove that (1) implies (2). Suppose p is irreducible and $p \mid ab$. If $a = 0$, then $p \mid a$. If $b = 0$, then $p \mid b$. Since p is not invertible, ab is not invertible. Write $ab = pc$ for some $c \in R$. Assume ab is nonzero and not invertible. Factor ab and pc into irreducibles. By uniqueness of factorization, p is an associate of one of the irreducible factors of a or b. □

4.2. Euclidean Domains.

DEFINITION 3.4.15. Let R be an integral domain. Then R is called a *euclidean domain* if there is a function (called the *norm*) $\delta : R - (0) \to \mathbb{N}$ such that

(1) $\delta(ab) = \delta(a)\delta(b)$ for all $a, b \in R - (0)$, and
(2) for all $a, b \in R - (0)$ there exist $q, r \in R$ such that $a = bq + r$ and either $r = 0$ or $\delta(r) < \delta(b)$.

EXAMPLE 3.4.16. The ring of integers $\mathbb{Z}$ is a euclidean domain with the norm function $\delta(x) = |x|$. The absolute value function is multiplicative, and property (2) is satisfied by the Division Algorithm on $\mathbb{Z}$.

EXAMPLE 3.4.17. We will prove above in Corollary 3.6.5 that if F is a field, then the polynomial ring $F[x]$ is a euclidean domain.

EXAMPLE 3.4.18. Let $n \geq 1$ and $\zeta_n = e^{2\pi i/n}$ a primitive nth root of unity in $\mathbb{C}$. Then $\mathbb{Q}[\zeta_n]$ is the splitting field for $x^n - 1$ over $\mathbb{Q}$ (see Example 5.2.9). Let $\mathbb{Z}[\zeta_n]$ be the subring of $\mathbb{Q}[\zeta_n]$ generated by adjoining ζ_n to $\mathbb{Z}$. So $\mathbb{Z}[\zeta_n]$ is the image of the evaluation homomorphism $\mathbb{Z}[x] \to \mathbb{C}$ defined by $x \mapsto \zeta_n$. When $n = 4$ we usually write i instead of ζ_4. In this case, the ring $\mathbb{Z}[i]$ is called the ring of *gaussian integers*.

EXAMPLE 3.4.19. In this example we prove that the ring of gaussian integers $\mathbb{Z}[i]$ (see Example 3.4.18) is a euclidean domain. Let $\chi : \mathbb{C} \to \mathbb{C}$ be complex conjugation: $\chi(a + bi) = a - bi$. The norm function $\delta : \mathbb{C} - (0) \to \mathbb{R}$ is defined by $\delta(a + bi) = a^2 + b^2 = (a + bi)\chi(a + bi)$. Since $\delta = 1_{\mathbb{C}}\chi$ is defined by multiplying two automorphisms, δ is multiplicative. Since $\min.\text{poly}(i) = x^2 + 1$, $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$. Now we prove that Property (2) of Definition 3.4.15 holds. Let $\alpha, \beta \in \mathbb{Z}[i] - (0)$. Since $\mathbb{Q}[i]$ is a field, we can

write $\alpha\beta^{-1} = u + vi$ where $u, v \in \mathbb{Q}$. Let $p, q \in \mathbb{Z}$ such that $|u-p| \leq 1/2$ and $|v-q| \leq 1/2$. Then $\gamma = p + qi \in \mathbb{Z}[i]$. Define $\rho = \alpha - \beta\gamma = \beta((u-p) + (v-q)i)$. Then

$$\begin{aligned}\delta(\rho) &= \delta\left(\beta((u-p)+(v-q)i)\right) \\ &= \delta(\beta)\left((u-p)^2 + (v-q)^2\right) \\ &\leq \delta(\beta)\left(\frac{1}{2^2} + (\frac{1}{2^2}\right) \\ &\leq \frac{1}{2}\delta(\beta) < \delta(\beta)\end{aligned}$$

and $\alpha = \beta\gamma + \rho$.

PROPOSITION 3.4.20. *If R is a euclidean domain, then R is a principal ideal domain.*

PROOF. Let I be a nonzero ideal in R. Consider the nonempty set $S = \{\delta(a) \mid a \in I - (0)\}$. By the Well Ordering Principle for $\mathbb{N}$, S has a least element, say $\delta(b)$, for some $b \in I$. Let $a \in I$. Since R is a euclidean domain, there exist q and r in R such that $a = bq + r$ and either $r = 0$ or $\delta(r) < \delta(b)$. Since $a, b \in I$, we have $r \in I$. By the minimal choice of $\delta(b)$, we conclude that $r = 0$. Thus, $a \in Rb$. This shows $I = Rb$ is principal. □

EXAMPLE 3.4.21. By Proposition 3.4.20, we have the following examples of principal ideal domains.

(1) We have not proved it yet, but if F is a field, then $F[x]$ is a principal ideal domain.
(2) By Example 3.4.19, the ring of gaussian integers $\mathbb{Z}[i]$ is a principal ideal domain.

PROPOSITION 3.4.22. *Let R be a euclidean domain with norm function $\delta : R - (0) \to \mathbb{N}$. Then the following are true:*

(1) $\delta(1) = 1$.
(2) If $u \in R^*$ is a unit in R, then $\delta(u) = 1$.
(3) If $\delta(u) = 1$, then $u \in R^*$ is a unit in R.
(4) The group of units of R is $R^* = \delta^{-1}\{1\}$.
(5) Let R be a euclidean domain with norm $\delta : R - (0) \to \mathbb{N}$. If $x \in R - (0)$ and $\delta(x) = 2$, then x is irreducible.

PROOF. (1) and (2): For any $u \in R - (0)$ we have $\delta(u) = \delta(u \cdot 1) = \delta(u)\delta(1)$. Therefore, $\delta(1) = 1$. Let $u \in R^*$. Then $1 = \delta(uu^{-1}) = \delta(u)\delta(u^{-1})$. Since the group of invertible elements of the ring $\mathbb{Z}$ is $\{1, -1\}$, we conclude that $\delta(u) = 1$.

(3): Assume $\delta(u) = 1$. Divide u into 1. There exist $q, r \in R$ such that $1 = uq + r$. Since 1 is the least element of $\mathbb{N}$, we conclude that $r = 0$. Thus, u is invertible.

(4): This part follows from (1), (2), and (3).

(5): Assume $x = ab$. Then $2 = \delta(x) = \delta(a)\delta(b)$. Thus $\delta(a) = 1$ or $\delta(b) = 1$. By (4), $R^* = \{u \in R - (0) \mid \delta(u) = 1\}$. Hence a is a unit or b is a unit. □

THEOREM 3.4.23. *If R is a euclidean domain with norm $\delta : R - (0) \to \mathbb{N}$, then R is a unique factorization domain.*

PROOF. The proof is in two parts. Part 1 proves the existence of the factorization, and Part 2 proves the uniqueness of the factorization.

(Existence.) Let $x \in R$ and assume $x \neq 0$ and x is not in R^*. Then $\delta(x) \geq 2$. The proof is by induction on $\delta(x)$. By Proposition 3.4.22, if $\delta(x) = 2$, then x is irreducible. This is the basis step. Inductively, assume $\delta(x) > 2$ and that if $y \in R - (0)$ and $1 < \delta(y) < \delta(x)$, then y has a factorization into irreducibles. If x is irreducible, then stop. Otherwise $x = x_1 x_2$

where x_1 and x_2 are both nonunits. Then $\delta(x) = \delta(x_1)\delta(x_2)$. We have $1 < \delta(x_i) < \delta(x)$ for $i = 1,2$. By our induction hypothesis, x_1 and x_2 can be represented as products of irreducibles. Therefore, x has such a factorization.

(Uniqueness.) Say $x = x_1 \cdots x_s = y_1 \cdots y_t$ are two representations of x as products of irreducibles. Since x_s is irreducible, by Corollary 3.4.10, there is some i such that $x_s \mid y_i$. Rearrange the factors if necessary, and assume $x_s \mid y_t$. Since y_t is irreducible, by Lemma 3.4.5 (5), x_s and y_t are associates. Cancel x_s and y_t. Then $x_1 \cdots x_{s-1}$ and $y_1 \cdots y_{t-1}$ are associates. By an inductive argument on the minimum of s and t, we see that $s = t$ and after rearranging if necessary, x_i and y_i are associates for each i. □

We end this section with a proof that in a euclidean domain R the greatest common divisor of two elements a and b can be computed by the Euclidean Algorithm (Proposition 3.4.24). In Corollary 3.4.25 we show that the same recursive algorithm also yields a solution (x,y) to the Bézout Identity $\gcd(a,b) = ax + by$.

PROPOSITION 3.4.24. *(The Euclidean Algorithm) Let R be a euclidean domain with with norm $\delta : R - (0) \to \mathbb{N}$. Let a and b be elements of R. The greatest common divisor of a and b exists and satisfies the following recursive formula:*

(1) (Basis) If $b = 0$, then $\gcd(a,b) = a$.
(2) (Recurrence) If $b \neq 0$, then $\gcd(a,b) = \gcd(b,r)$, where $a = bq + r$ and either $r = 0$ or $\delta(r) < \delta(b)$.

PROOF. If $b = 0$, then the ideals (a,b) and (a) are equal in R, and Corollary 3.4.9 implies $\gcd(a,b) = a$. If $b \neq 0$, then by Definition 3.4.15, $a = bq + r$, for elements q and r in R such that either $r = 0$ or $\delta(r) < \delta(b)$. Then the ideals (a,b) and (b,r) are equal in R. By Corollary 3.4.9, $\gcd(a,b) = \gcd(b,r)$. To see that the recursive algorithm converges, set $r_0 = b$ and successively apply Definition 3.4.15 to find a sequence of quotients $q_1, q_2, \ldots, q_{n+1}$ and a sequence of remainders $r_0, r_1, r_2, \ldots, r_n$ satisfying:

$$\begin{aligned}
a &= r_0 q_1 + r_1, \quad 0 < \delta(r_1) < \delta(r_0) \\
r_0 &= r_1 q_2 + r_2, \quad 0 < \delta(r_2) < \delta(r_1) \\
r_1 &= r_2 q_3 + r_3, \quad 0 < \delta(r_3) < \delta(r_2) \\
&\vdots \\
r_{n-3} &= r_{n-2} q_{n-1} + r_{n-1}, \quad 0 < \delta(r_{n-1}) < \delta(r_{n-2}) \\
r_{n-2} &= r_{n-1} q_n + r_n, \quad 0 < \delta(r_n) < \delta(r_{n-1}) \\
r_{n-1} &= r_n q_{n+1} + 0
\end{aligned}$$

where r_n is the last nonzero remainder. The algorithm converges for some n such that $0 \leq n \leq \delta(b)$ because $\delta(r_0) > \delta(r_1) > \delta(r_2) > \cdots > \delta(r_n) > 0$. As mentioned above,

$$\begin{aligned}
r_n = \gcd(r_n, r_{n-1}) = \gcd(r_n, r_{n-1}) = \gcd(r_{n-1}, r_{n-2}) \\
= \cdots = \gcd(r_3, r_2) = \gcd(r_2, r_1) = \gcd(r_1, r_0) = \gcd(a,b).
\end{aligned}$$

□

COROLLARY 3.4.25. *(Bézout's Identity) Let R be a euclidean domain with norm function $\delta : R - (0) \to \mathbb{N}$. Let a and b be elements of R. There exist x, y in R such that $\gcd(a,b) = ax + by$.*

PROOF. If $a = 0$, then $b = \gcd(a,b)$. Take $x = 0$ and $y = 1$. If $b = 0$, then $a = \gcd(a,b)$. Take $x = 1$ and $y = 0$. If $b \neq 0$, then by Definition 3.4.15, $a = bq + r$, for elements q and r

in R such that either $r = 0$ or $\delta(r) < \delta(b)$. Then $\gcd(a,b) = \gcd(b,r)$ and by induction on $\delta(b)$ we can write $\gcd(b,r) = bu + rv$ for some u, v in R. Then

$$\begin{aligned}\gcd(a,b) &= bu + rv \\ &= bu + (a - bq)v \\ &= av + b(u - qv).\end{aligned}$$

Take $x = v$ and $y = u - qv$. □

4.3. Principal Ideal Domains. The fundamental properties of a principal ideal domain are derived in Theorem 3.4.26. In particular, every principal ideal domain is a unique factorization domain. Part (2) shows that a PID satisfies the ascending chain condition on ideals. An integral domain with this property is said to be *noetherian*.

THEOREM 3.4.26. *(Fundamental Theorem on Principal Ideal Domains) Let R be a principal ideal domain (a PID, for short).*

(1) If p is an irreducible element, then p is a prime element.
(2) R satisfies the ascending chain condition *on ideals. That is, given a chain of ideals $I_1 \subseteq I_2 \subseteq I_3 \subseteq \cdots \subseteq I_n \subseteq \cdots$, there exists $N \geq 1$ such that $I_N = I_{N+1} = \cdots$.*
(3) If $a \in R$ is a nonunit, nonzero element of R, then the set

$$\mathscr{S} = \{p \in R \mid p \text{ is irreducible and } p \mid a\}$$

contains only a finite number of associate classes. In other words, up to associates, a has only a finite number of irreducible factors.
(4) If I is an ideal in R which is not the unit ideal, then $\bigcap_{n\geq 1} I^n = (0)$.
(5) Suppose a is a nonzero element in R, p is irreducible and p is a factor of a. Then for some $n \geq 1$ we have $a \in (p^n)$ and $a \notin (p^{n+1})$.
(6) If $a \in R$ is a nonunit and a nonzero element, then there exists an irreducible element p such that $p \mid a$.
(7) R is a unique factorization domain.

PROOF. (1): This is Corollary 3.4.10.

(2): Let $I = \bigcup_{k=1}^{\infty} I_k$. By Exercise 3.2.34, I is an ideal in R. Since R is a PID, there exists $a \in R$ such that $I = (a)$. Given $a \in I$, we know $a \in I_N$ for some N. Then $I = (a) \subseteq I_N \subseteq I_{N+1} \subseteq \cdots$ and we are done.

(3): The proof is by contradiction. Assume $\{p_1, p_2, \dots\}$ is a sequence in $\mathscr{S}$ such that for each $n > 1$, p_n does not divide $p_1 p_2 \cdots p_{n-1}$. Write $a = p_1 a_1$. Then $p_2 \mid p_1 a_1$. By assumption, p_2 does not divide p_1. By Part (1), $p_2 \mid a_1$ and we write $a_1 = p_2 a_2$. Iteratively we arrive at the factorizations

$$a = p_1 a_1 = p_1 p_2 a_2 = \cdots = p_1 p_2 \cdots p_n a_n.$$

Applying one more step, we know $p_{n+1} \mid a$. Since p_{n+1} does not divide $p_1 p_2 \cdots p_n$, and p_{n+1} is prime, it follows that $p_{n+1} \mid a_n$. Write $a_n = p_{n+1} a_{n+1}$. Therefore $(a_n) \subseteq (a_{n+1})$ with equality if and only if a_n and a_{n+1} are associates. But p_{n+1} is not a unit, so by Lemma 3.4.3 (4), the chain of ideals

$$(a_1) \subseteq (a_2) \subseteq \cdots \subseteq (a_n) \subseteq (a_{n+1}) \subseteq \cdots$$

is strictly increasing. This contradicts Part (2).

(4): Because R is a PID, $I = (b)$ for some $b \in R$. If $I = 0$, then Part (4) is trivial, so we assume $b \neq 0$. Let $M = \bigcap_{n=1}^{\infty} I^n$. Then M is an ideal in R, so $M = (r)$ for some $r \in R$. Since M is an ideal, $bM \subseteq M$. To show that $bM = M$, assume $x \in M$. Then

$x \in M \subseteq I$ implies $x = by$ for some $y \in R$. Let $n \geq 1$. Then $x \in M \subseteq I^{n+1} = (b^{n+1})$ implies $x = b^{n+1}z$ for some $z \in R$. Since R is an integral domain and $b \neq 0$, $x = by = b^{n+1}z$ implies $y = b^n z \in I^n = (b^n)$. This proves $y \in \bigcap_{n\geq 1} I^n = M$. Therefore $x \in bM$, and $bM = M$. Since $bM = (br)$, Lemma 3.4.3 says br and r are associates. But b is not a unit, so $r = 0$, which proves (4).

(5): Set $I = (p)$. By assumption, $a \in (p)$ and $a \neq 0$. By Part (4), for some $n \geq 1$ we have $a \notin (p^{n+1})$ and $a \in (p^n)$.

(6): The proof is by contradiction. Suppose $a \in R$ is not a unit, and not divisible by an irreducible. Then a is not irreducible. There are nonunits a_1, b_1 in R such that $a = a_1 b_1$. By our assumption, a_1 and b_1 are not irreducible. By Lemma 3.4.3, $(a) \subsetneq (a_1)$. Since a_1 is not irreducible, there are nonunits a_2, b_2 in R such that $a_1 = a_2 b_2$. Since a_2 and b_2 are divisors of a, both are not irreducible. By Lemma 3.4.3, $(a) \subsetneq (a_1) \subsetneq (a_2)$. Recursively we construct a strictly increasing sequence of ideals $(a_i) \subsetneq (a_{i+1})$, contradicting Part (2).

(7): This proof is left to the reader. □

4.4. Exercises.

EXERCISE 3.4.27. Let a and b be elements of a commutative ring R. If $(a,b) = (1)$ and $a \mid bc$, then $a \mid c$.

EXERCISE 3.4.28. Let X be a nonempty subset of a commutative ring R. If $d \in (X)$ and $d \mid x$ for all $x \in X$, then $(d) = (X)$.

EXERCISE 3.4.29. Let $X = \{x_1, \ldots, x_n\}$ be a nonempty finite subset of a commutative ring R, with $n \geq 2$. If $e = \gcd(x_1, \ldots, x_{n-1})$ and $d = \gcd(e, x_n)$, then $d = \gcd(x_1, \ldots, x_n)$.

EXERCISE 3.4.30. (Exponential Notation in a UFD) Let a and b be elements of a unique factorization domain R. Assume a and b are both nonzero and nonunits.

(1) Show that there exist irreducible elements $x_1, \ldots, x_m$ in R such that x_i and x_j are associates of each other if and only if $i = j$ and nonnegative integers $e_1, \ldots, e_m$, $f_1, \ldots, f_m$ such that $a = x_1^{e_1} \cdots x_m^{e_m}$ and $b = x_1^{f_1} \cdots x_m^{f_m}$.
(2) Show that in the notation from (1) that $a \mid b$ if and only if $e_i \leq f_i$ for each i.
(3) In the notation from (1), for $j = 1, \ldots, m$, let ℓ_j be the least element in the set $\{e_j, f_j\}$. Prove that $d = x_1^{\ell_1} x_2^{\ell_2} \cdots x_m^{\ell_m} = \gcd(a,b)$.

EXERCISE 3.4.31. Let R be an integral domain and X a nonempty subset of R. Assume $d = \gcd(X)$ exists and $d \neq 0$. Let $Y = \{xd^{-1} \mid x \in X\}$ (see Lemma 3.4.3 (5) for this notation). Prove that $1 = \gcd(Y)$.

5. The Quotient Field of an Integral Domain

Let R be an integral domain. Define a relation on $R \times (R - (0))$ by the rule: $(r,v) \sim (s,w)$ if and only if $rw = sv$. We show that $\sim$ is an equivalence relation. Clearly $\sim$ is reflexive and symmetric. Let us show that it is transitive. Suppose $(r,u) \sim (s,v)$ and $(s,v) \sim (t,w)$. Then $rv = su$ $sw = tv$. Multiply the first by w and the second by u to get $rvw = suw = tvu$. Then $rvw = tvu$. Canceling v, $rw = tu$, which implies $(r,u) \sim (t,w)$. We have shown that $\sim$ is an equivalence relation on $R \times W$. The set of equivalence classes, $(R \times (R - (0))) / \sim$, is called the *quotient field*, or *field of fractions of* R. The equivalence class containing (r,w) is denoted by the fraction r/w.

LEMMA 3.5.1. *Let R be an integral domain and $K = (R \times (R-(0)))/\sim$ the quotient field of R. Then K is a field with the binary operations*

$$\frac{r}{v} + \frac{s}{w} = \frac{rw+sv}{vw}, \quad \frac{r}{v}\frac{s}{w} = \frac{rs}{vw}.$$

The additive identity is $0/1$, the multiplicative identity is $1/1$. There is a natural map $\theta : R \to K$ defined by $r \mapsto r/1$ which is a one-to-one homomorphism of rings. If R is a field, then θ is an isomorphism.

PROOF. Assume $\frac{r}{v} = \frac{r_1}{v_1}$ and $\frac{s}{w} = \frac{s_1}{w_1}$. Then

$$rv_1 = r_1 v \tag{5.1}$$

$$sw_1 = s_1 w. \tag{5.2}$$

Multiply (5.1) by ww_1 and (5.2) by vv_1 to get the identities $rv_1ww_1 = r_1vww_1$ and $sw_1vv_1 = s_1wvv_1$. From these we derive

$$\begin{aligned}(rw+sv)v_1w_1 &= rv_1ww_1 + sw_1vv_1 \\ &= r_1vww_1 + s_1wvv_1 \\ &= (r_1w_1 + s_1v_1)vw.\end{aligned}$$

This is the center equation in:

$$\frac{r}{v} + \frac{s}{w} = \frac{rw+sv}{vw} = \frac{r_1w_1+s_1v_1}{v_1w_1} = \frac{r_1}{v_1} + \frac{s_1}{w_1}.$$

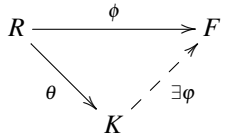
Hence, addition of fractions is well defined. Multiply (5.1) by sw_1 and (5.2) by r_1v to get the identities $rsv_1w_1 = r_1vsw_1$ and $sw_1r_1v = s_1wr_1v$. Taken together, we have $rsv_1w_1 = r_1vsw_1 = s_1wr_1v$. This is the center equation in:

$$\frac{r}{v}\frac{s}{w} = \frac{rs}{vw} = \frac{r_1s_1}{v_1w_1} = \frac{r_1}{v_1}\frac{s_1}{w_1}.$$

Hence, multiplication of fractions is well defined. It is routine to check that the associative and distributive laws hold, that K is a field, and that θ is a one-to-one homomorphism of rings. The details are left to the reader. □

5.1. Exercises.

EXERCISE 3.5.2. (Universal Mapping Property) Let R be an integral domain with field of fractions K. Let F be a field and $\phi : R \to F$ a one-to-one homomorphism of rings. Prove that there is a unique homomorphism of fields $\varphi : K \to F$ such that the diagram



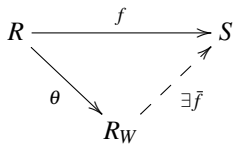
commutes where θ is the natural map of Lemma 3.5.1.

EXERCISE 3.5.3. Let R be a commutative ring. A subset W of R is called a *multiplicative subset of R*, if the following three properties hold:

(a) $1 \in W$.
(b) W contains no zero divisor of R.
(c) If x and y are in W, then $xy \in W$.

If W is a multiplicative subset of R, do the following:

(1) Define a relation on $R \times W$ by the rule: $(r,v) \sim (s,w)$ if and only if $rw = sv$. Show that $\sim$ is an equivalence relation. Denote the set of equivalence classes by R_W.
(2) Show how to make R_W into a commutative ring by imitating the construction of the quotient field of an integral domain in Lemma 3.5.1. The ring R_W is called the *localization* of R at W.
(3) Show that there is a one-to-one homomorphism of rings $\theta : R \to R_W$.
(4) (Universal Mapping Property) Let S be a commutative ring and $f : R \to S$ a homomorphism such that $f(W) \subseteq \text{Units}(S)$. Show that there exists a unique homomorphism $\bar{f} : R_W \to S$



such that $f = \bar{f}\theta$.

EXERCISE 3.5.4. Let R be a commutative ring and W the set of all elements of R that are not zero divisors.

(1) Show that W is a multiplicative subset. In this case, the localization R_W is called the *total ring of quotients* of R.
(2) Let S be the total ring of quotients of R. Show that S is a commutative ring with the property that every element of S is either a unit or a zero divisor.

EXERCISE 3.5.5. Let R be a finite ring in which $0 \neq 1$, and $x \in R$. Show that if x is not a zero divisor, then x is invertible. (Hint: Theorem 3.2.20.)

EXERCISE 3.5.6. Recall that $\mathbb{Z}[\sqrt{-5}]$ is the subring of $\mathbb{Q}[\sqrt{-5}]$ generated by $\mathbb{Z}$ and $\sqrt{-5}$. Show that the field of fractions of $\mathbb{Z}[\sqrt{-5}]$ is $\mathbb{Q}[\sqrt{-5}]$.

6. Polynomial Rings

Let R be a commutative ring. The *polynomial ring in one variable x with coefficients in R*,

$$R[x] = \left\{ \sum_{i=0}^{n} a_i x^i \mid n \geq 0, a_i \in R \right\}$$

is constructed in the usual way. It is assumed that the *indeterminate* x commutes with elements of R. The ring $R[x]$ is commutative. If $a \in R - (0)$, the *degree* of the *monomial* ax^n is n. For convenience, the degree of 0 is taken to be $-\infty$. The *degree* of a polynomial $f = \sum_{i=0}^{n} a_i x^i$ in $R[x]$ is the maximum of the degrees of the terms $a_0 x^0, \ldots, a_n x^n$. If f is nonzero of degree n, the *leading coefficient* of f is a_n. We say that f is *monic* if the leading coefficient of f is 1. If $f = \sum_{i=0}^{m} a_i x^i$ has degree m and $g = \sum_{i=0}^{n} b_i x^i$ has degree n, then

$$\begin{aligned} fg &= \left(\sum_{i=0}^{m} a_i x^i \right) \left(\sum_{i=0}^{n} b_i x^i \right) \\ &= a_0 b_0 + (a_0 b_1 + a_1 b_0)x + \cdots + \left(\sum_{j=0}^{k} a_j b_{k-j} \right) x^k + \cdots + a_m b_n x^{m+n}. \end{aligned}$$

It follows that $\deg(fg) = \deg(f) + \deg(g)$ in case one of the leading coefficients a_m or b_n is not a zero divisor in R. The degree of a sum is no larger than the degree of either summand: $\deg(f+g) \leq \max(\deg(f), \deg(g))$. We view R as the subring of all polynomials in $R[x]$ of degree less than or equal to 0. The natural mapping $R \to R[x]$ which maps $a \in R - (0)$ to the polynomial of degree zero is a monomorphism. The polynomial ring over R in several variables is defined by iterating the one-variable construction. If $t > 1$ and $x_1, \ldots, x_t$ are indeterminates, then $R[x_1, \ldots, x_t] = R[x_1, \ldots, x_{t-1}][x_t]$. See Section 3.6.1.

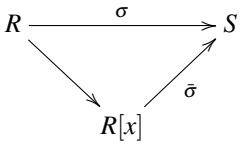
PROPOSITION 3.6.1. *If R is an integral domain if and only if $R[x]$ is an integral domain. In general, R is an integral domain if and only if $R[x_1, \ldots, x_t]$ is an integral domain.*

PROOF. Since R is a subring of $R[x]$, if R has a nonzero zero divisor, so does $R[x]$. Suppose R is an integral domain. Let f and g be nonzero polynomials in $R[x]$. Say $f = \sum_{i=0}^{m} a_i x^i$ has degree m and $g = \sum_{i=0}^{n} b_i x^i$ has degree n. Then the leading term of fg is $a_m b_n \neq 0$. This proves $R[x]$ is an integral domain. If $t > 1$, the proof follows by Mathematical Induction. □

THEOREM 3.6.2. *Let R be a commutative ring and $\sigma : R \to S$ a homomorphism of rings.*

(1) If S is a commutative ring, the definition $\bar{\sigma}(\sum r_i x^i) = \sum \sigma(r_i) x^i$ extends σ to a homomorphism on the polynomial rings $\bar{\sigma} : R[x] \to S[x]$. If $K = \ker(\sigma)$, then the kernel of $\bar{\sigma}$ is the set $K[x]$ consisting of those polynomials $f \in R[x]$ such that every coefficient of f is in K.

(2) (Universal Mapping Property) Let s be an element of S such that $s\sigma(r) = \sigma(r)s$ for every $r \in R$. Then there is a unique homomorphism $\bar{\sigma}$ such that $\bar{\sigma}(x) = s$ and the diagram



commutes. We say $\bar{\sigma}$ is the evaluation homomorphism *defined by $x \mapsto s$.*

PROOF. The proof is left to the reader. □

THEOREM 3.6.3. *(The Division Algorithm) Let R be a commutative ring. Let $f, g \in R[x]$ and assume the leading coefficient of g is a unit of R. There exist unique polynomials $q, r \in R[x]$ such that $f = qg + r$ and $\deg r < \deg g$.*

PROOF. (Existence.) If $\deg f < \deg g$, then set $q = 0$ and $r = f$. Otherwise assume $f = \sum_{i=0}^{m} a_i x^i$ where $a_m \neq 0$ and $g = \sum_{i=0}^{n} b_i x^i$ where $b_n \neq 0$ and b_n is a unit in R. If $m = 0$, then $n = 0$ so $q = a_0 b_0^{-1}$ and $r = 0$. Proceed by induction on m. The leading coefficient of $\left(a_m b_n^{-1} x^{m-n}\right) g$ is a_m. Set $h = f - \left(a_m b_n^{-1} x^{m-n}\right) g$. Then $\deg h < \deg f$. By induction, $h = q_1 g + r$ where $\deg r < \deg g$. Now

$$\begin{aligned} f &= \left(a_m b_n^{-1} x^{m-n}\right) g + q_1 g + r \\ &= \left(a_m b_n^{-1} x^{m-n} + q_1\right) g + r \end{aligned}$$

so take $q = a_m b_n^{-1} x^{m-n} + q_1$.

(Uniqueness.) Assume $f = gq + r = gq_1 + r_1$ where $\deg r < \deg g$ and $\deg r_1 < \deg g$. Subtracting, we have $g(q - q_1) = r_1 - r$. The degree of the right hand side is $\deg(r_1 - r) \leq \max(\deg r_1, \deg r) < \deg g$. The degree of the left hand side is $\deg g + \deg(q - q_1)$. If

$q - q_1 \neq 0$, this is impossible. So $q_1 = q$ and $r = r_1$. Hence the quotient and remainder are unique. □

COROLLARY 3.6.4. *(Synthetic Division) If R is a commutative ring, $f = \sum_{i=0}^{m} r_i x^i \in R[x]$ and $a \in R$, then there exists a unique polynomial $q \in R[x]$ such that $f = q(x-a) + f(a)$ where $f(a) = \sum_{i=0}^{m} r_i a^i \in R$.*

PROOF. Upon dividing $x - a$ into f, this follows straight from Theorem 3.6.3. □

COROLLARY 3.6.5. *If k is a field, then $k[x]$ is a euclidean domain. It follows that $k[x]$ is a PID and a UFD.*

PROOF. Define the norm function by the exponential formula: $\delta(f) = 2^{\deg f}$ for all $f \in F[x] - (0)$. Then $\delta(fg) = 2^{\deg fg} = 2^{\deg f + \deg g} = 2^{\deg f} 2^{\deg g} = \delta(f)\delta(g)$, hence δ is multiplicative. In Definition 3.4.15, property (2) is the division algorithm on $F[x]$. □

If k is a field, and $R = k[x]$, then the quotient field of $k[x]$, denoted $k(x)$, is called the field of rational functions over k. If S is a ring and R a subring, then by Theorem 3.6.2 we can view $R[x]$ as a subring of $S[x]$.

DEFINITION 3.6.6. Let R be a commutative ring, $u \in R$, and $f = \sum_{i=0}^{m} r_i x^i \in R[x]$. We say that u is a *root* of f in case $f(u) = \sum_{i=0}^{m} r_i u^i = 0$.

LEMMA 3.6.7. *Let R be a commutative ring, $u \in R$, and $f \in R[x]$. The following are equivalent.*

(1) u is a root of f.
(2) f is in the kernel of the evaluation homomorphism $R[x] \to R$ defined by $x \mapsto u$.
(3) There exists $q \in R[x]$ such that $f = (x-u)q$.

PROOF. The proof is left to the reader. □

COROLLARY 3.6.8. *If R is an integral domain, and $f \in R[x]$ has degree $d \geq 0$, then the following are true:*

(1) If u is a root of f in R, then there exists $m \geq 1$ such that $f = (x-u)^m q$ and $q(u) \neq 0$.
(2) f has at most d roots in R.
(3) (Lagrange Interpolation) Let $n \geq 1$. Given $n+1$ distinct elements of R: $\alpha_0, \ldots, \alpha_n$, and $n+1$ arbitrary elements of R: $\beta_0, \ldots, \beta_n$, there exists a unique polynomial $f \in R[x]$ such that $\deg f \leq n$ and $f(\alpha_i) = \beta_i$ for each i.

PROOF. (1): Apply Lemma 3.6.7 and induction on the degree.
(2): If $d = 0$, then f has no root. Inductively assume $d \geq 1$ and that the result holds for any polynomial of degree in the range $0, \ldots, d-1$. If f has no root, then we are done. Suppose u is a root of f. By Part (1) we can write $f = (x-u)^m q$, where $\deg q = d - m$. If $v \neq u$ is another root of f, then $0 = f(u) = (v-u)^m q(u)$. Since R is an integral domain, this means u is a root of q. By induction, there are at most $d - m$ choices for v.

(3): (Existence.) The Lagrange basis polynomials with respect to the set $\{\alpha_0, \ldots, \alpha_n\}$ are

$$L_0(x) = \frac{(x-\alpha_1)\cdots(x-\alpha_n)}{(\alpha_0-\alpha_1)\cdots(\alpha_0-\alpha_n)}$$

$$\vdots$$

$$L_j(x) = \frac{(x-\alpha_0)\cdots(x-\alpha_{j-1})(x-\alpha_{j+1})\cdots(x-\alpha_n)}{(\alpha_j-\alpha_0)\cdots(\alpha_j-\alpha_{j-1})(\alpha_j-\alpha_{j+1})\cdots(\alpha_j-\alpha_n)}$$

$$\vdots$$

$$L_n(x) = \frac{(x-\alpha_0)\cdots(x-\alpha_{n-1})}{(\alpha_n-\alpha_0)\cdots(\alpha_n-\alpha_{n-1})}.$$

Notice that $L_j(x)$ has degree n and

$$L_j(\alpha_k) = \begin{cases} 0 & \text{if } k \neq j \\ 1 & \text{if } k = j. \end{cases}$$

Set

$$f(x) = \sum_{j=0}^{n} \beta_j L_j(x).$$

Then $f(\alpha_k) = \beta_k$ for each $k = 0, \ldots, n$ and $\deg f \leq n$.

(Uniqueness.) Suppose f and g are two polynomials in $R[x]$ such that $\deg f \leq n$, $\deg g \leq n$ and $f(\alpha_k) = \beta_k = g(\alpha_k)$ for each $k = 0, \ldots, n$. Then $\deg(f-g) \leq n$ and $f-g$ has $n+1$ roots, namely $\alpha_0, \ldots, \alpha_n$. By Part (2), $f-g = 0$. □

COROLLARY 3.6.9. *Let R be an integral domain. Let $n > 1$ be an integer. The group of nth roots of unity in R, $\mu_n = \{u \in R \mid u^n = 1\}$, is a cyclic group of order at most n.*

PROOF. The set μ_n is clearly a subgroup of R^*. The order of μ_n is at most n, by Corollary 3.6.8 (2). For every divisor d of n, the equation $x^d = 1$ has at most d solutions in R^*. By Exercise 2.8.10, μ_n is a cyclic group. □

COROLLARY 3.6.10. *Let F be a finite field of order q. Then F^* is a cyclic abelian group of order $q-1$.*

PROOF. In a field the nonzero elements make up an abelian group. The group F^* has order $q-1$. By Corollary 2.2.17, every $u \in F^*$ satisfies the equation $u^{q-1} = 1$. By Corollary 3.6.9, F^* is a cyclic group of order $q-1$. □

EXAMPLE 3.6.11. If F is a field, the ring $F[x,y]$ is not a PID. The ideal $(x,y) = \{ux + vy \mid u, v \in F[x,y]\}$ is not a principal ideal.

DEFINITION 3.6.12. If R is an integral domain, $f \in R[x]$, and u is a root of f, then the *multiplicity* of u as a root of f is the positive number m given by Corollary 3.6.8 (1). We say that u is a *simple root* if $m = 1$. If $m > 1$, then u is called a *multiple root*.

DEFINITION 3.6.13. If R is any ring and $f = \sum_{i=0}^{n} a_i x^i \in R[x]$, then the *formal derivative* of f is defined to be

$$f' = \sum_{i=1}^{n} i a_i x^{i-1}$$

which is also in $R[x]$. The reader should verify the usual identities satisfied by the derivative operator. In particular, $(af+bg)' = af'+bg'$ and $(fg)' = f'g+fg'$. If R is commutative, then $(f^n)' = nf^{n-1}f'$.

PROPOSITION 3.6.14. *Suppose S is an integral domain and R is a subring of S. Let f be a nonconstant polynomial in $R[x]$ and $u \in S$. Then u is a multiple root of f if and only if $f'(u) = f(u) = 0$.*

PROOF. Suppose u is a multiple root of f. Write $f = (x-u)^2 q$ for some $q \in S[x]$ and compute $f' = 2(x-u)q + (x-u)^2 q'$. It is immediate that $f'(u) = 0$. Conversely, assume $f(u) = f'(u) = 0$. Write $f = (x-u)q$ for some $q \in S[x]$ and compute $f' = q + (x-u)q'$. It is immediate that $q(u) = 0$, so $f = (x-u)^2 q_2$ for some $q_2 \in S[x]$. □

THEOREM 3.6.15. *Let k be a subfield of the integral domain S and f a nonconstant polynomial in $k[x]$.*

(1) Assume
  (a) $\gcd(f, f') = 1$, or
  (b) f is irreducible in $k[x]$ and $f' \neq 0$ in $k[x]$, or
  (c) f is irreducible in $k[x]$ and k has characteristic zero (see Example 3.2.4 (5)).
  Then f has no multiple root in S.

(2) Suppose p denotes the characteristic of k. Assume u is a root of f in S.
  (a) If f is irreducible in $k[x]$ and u is a multiple root of f, then $p > 0$ and $f \in k[x^p]$.
  (b) If $p > 0$ and $f \in k[x^p]$, then u is a multiple root of f.

PROOF. (1): Assuming $\gcd(f, f') = 1$, by Corollary 3.4.9 there exist $s, t \in k[x]$ such that $1 = fs + f't$. It is clear that f and f' do not have a common root in S. By Proposition 3.6.14, f has no multiple root in S. Case (b) reduces immediately to case (a). Case (c) reduces immediately to case (b).

(2) (a): If $u \in S$ is a multiple root of f, then because f is irreducible in $k[x]$, Part (1) implies $p > 0$ and $f' = 0$. The reader should verify that under these conditions $f \in k[x^p]$.

(2) (b): If k has characteristic $p > 0$ and $f \in k[x^p]$, then clearly $f' = 0$. If $u \in S$ is a root of f, then by Proposition 3.6.14, u is a multiple root of f. □

6.1. Polynomials in Several Variables. The polynomial ring over R in several variables is defined by iterating the one-variable construction. If $m > 1$ and $x_1, \ldots, x_m$ are indeterminates, then $R[x_1, \ldots, x_m] = R[x_1, \ldots, x_{m-1}][x_m]$. A *monomial* in $S = R[x_1, \ldots, x_m]$ is a polynomial of the form $M = r x_1^{e_1} \cdots x_m^{e_m}$, where $r \in R$ is the *coefficient* and each exponent e_i is a nonnegative integer. The *degree* of a monomial is $-\infty$ if $r = 0$, otherwise it is the sum of the exponents. If $M \neq 0$, then $\deg M = e_1 + \cdots + e_m$. If M_1 and M_2 are monomials with coefficients r_1, r_2, then $M_1 M_2$ is a monomial with coefficient $r_1 r_2$. So $M_1 M_2 = 0$ if and only if $r_1 r_2 = 0$. If $M_1 M_2 \neq 0$, then $\deg M_1 M_2 = \deg M_1 + \deg M_2$. A polynomial f in S is a sum $f = \sum_{j=1}^{d} M_j$ where each M_j is a monomial. A polynomial $f \in S$ is said to be *homogeneous* if f can be written as a sum of monomials all of the same degree. Let $S_0 = R$ be the set of all polynomials in S of degree less than or equal to 0. For all $n \geq 1$, let S_n be the R-submodule generated by the set of all homogeneous polynomials in S of degree n. If f is homogeneous of degree d and g is homogeneous of degree e, then we see fg is homogeneous of degree $d+e$. A polynomial $f \in S$ can be written $f = f_0 + f_1 + \cdots + f_d$ where each f_i is homogeneous of degree i. We call f_i the *homogeneous component of f of degree i*. This representation of f as a sum of homogeneous polynomials is unique. The *degree* of a polynomial is the maximum of the degrees of the homogeneous components.

If k is a field, then $k[x_1,\dots,x_m]$ is an integral domain. The quotient field of $k[x_1,\dots,x_m]$, denoted $k(x_1,\dots,x_m)$, is called the field of rational functions over k in m variables.

In Exercise 1.2.23 the lexicographical order $\leq$ is defined on the set of all m-tuples of nonnegative integers $\prod_{i=1}^m \mathbb{Z}_{\geq 0} = \{(e_1,\dots,e_m) \mid x_i \in \mathbb{Z}_{\geq 0}\}$. Under this partial ordering $\prod_{i=1}^m \mathbb{Z}_{\geq 0}$ is a chain. This notion induces the *lexicographical order* on the set of nonzero monomials in $R[x_1,\dots,x_m]$. If $M_1 = r_1 x_1^{a_1}\cdots x_m^{a_m}$, and $M_2 = r_2 x_1^{b_1}\cdots x_m^{b_m}$ are two nonzero monomials, then $M_1 < M_2$ if and only if $(a_1,\dots,a_m) < (b_1,\dots,b_m)$. We see that M_1 and M_2 are comparable if $(a_1,\dots,a_m) \neq (b_1,\dots,b_m)$.

LEMMA 3.6.16. *Let R be a ring and $S = R[x_1,\dots,x_m]$.*

(1) A nonzero polynomial f in S can be written as a sum $f = \sum_{j=1}^d M_j$ where each M_j is a nonzero monomial such that $M_1 < M_2 < \cdots < M_d$. This representation as a sum of strictly increasing monomials is unique. The monomial M_d is called the leading term *of f.*

(2) Let f and g be nonzero polynomials in S. Let $L(f)$ be the leading term of f and $L(g)$ the leading term of g. Then the leading term of fg is equal to $L(f)L(g)$.

(3) If U is a nonempty set of nonzero monomials in S, then there exists an element $\alpha \in U$ with the property that if $\beta \in U$ and β is comparable to α, then $\alpha < \beta$. If U has the property that any two distinct elements are comparable, then there exists $\alpha \in U$ such that if $\beta \in U - \{\alpha\}$, then $\alpha < \beta$.

PROOF. (1): Given a nonzero polynomial f, write $f = \sum_{j=1}^d M_j$ where each M_j is a nonzero monomial. By adding coefficients, all monomials that are incomparable can be combined. Hence we can assume the monomials appearing in the sum are comparable. After rearranging if necessary, we can assume $M_1 < M_2 < \cdots < M_d$. Conversely, if $M_1 < M_2 < \cdots < M_d$ is a strictly increasing sequence of monomials, then the sum $f = \sum_{j=1}^d M_j$ is nonzero. The uniqueness claim follows from this fact.

(2): The proof of this part is left to the reader.

(3): This follows from Exercise 1.2.23 (3). □

6.2. Exercises.

EXERCISE 3.6.17. Let k be a field. Let $R = k[x^2, x^3]$ be the subring of $k[x]$ consisting of all polynomials such that the coefficient of x is zero. Prove:

(1) R is an integral domain.
(2) R is not a UFD. (Hint: x^2 and x^3 are both irreducible.)
(3) R is not a PID. (Hint: Neither x^2 nor x^3 is prime.)
(4) The converse of Lemma 3.4.5 (2) is false.

EXERCISE 3.6.18. Let R be a commutative ring and $I = (a)$ a principal ideal in R. Show that for any $n \geq 1$, $I^n = (a^n)$.

EXERCISE 3.6.19. Prove that if R is an integral domain, then the homomorphism $R \to R[x]$ induces an isomorphism on the groups of units $\mathrm{Units}(R) \to \mathrm{Units}(R[x])$.

EXERCISE 3.6.20. Let R be a commutative ring. Prove:

(1) The nil radical of $R[x]$ is equal to $\mathrm{Rad}_R(0)[x]$. That is, a polynomial is nilpotent if and only if every coefficient is nilpotent.
(2) The kernel of $R[x] \to (R/\mathrm{Rad}_R(0))[x]$ is equal to the nil radical of $R[x]$.
(3) The group of units of $R[x]$ consists of those polynomials of the form $f = a_0 + a_1 x + \cdots + a_n x^n$, where a_0 is a unit in R and $f - a_0 \in \mathrm{Rad}_R(0)[x]$.

(4) If $\mathrm{Rad}_R(0) = (0)$, then the homomorphism $R \to R[x]$ induces an isomorphism on the groups of units $\mathrm{Units}(R) \to \mathrm{Units}(R[x])$.

EXERCISE 3.6.21. Let R be an integral domain and $a \in R$. Prove that the linear polynomial $x - a$ is a prime element in $R[x]$.

EXERCISE 3.6.22. Let R be a commutative ring and $a \in R$. Show that there is an automorphism $\theta : R[x] \to R[x]$ such that $\theta(x) = x + a$ and for all $r \in R$, $\theta(r) = r$.

EXERCISE 3.6.23. Let R be an integral domain and a an irreducible element of R. Prove that a is an irreducible element in $R[x]$.

EXERCISE 3.6.24. Let k be a field and $A = k[x]$. Prove:

(1) If $I = (x)$ is the ideal in A generated by x, then $I^n = (x^n)$.
(2) Let $n \geq 1$. The nil radical of $k[x]/(x^n)$ consists of those cosets represented by polynomials of the form $\alpha_1 x + \cdots + \alpha_{n-1}x^{n-1}$.
(3) The group of units of $k[x]/(x^n)$ consists of those cosets represented by polynomials of the form $\alpha_0 + \alpha_1 x + \cdots + \alpha_{n-1}x^{n-1}$, where α_0 is a unit in k.

EXERCISE 3.6.25. Let R be an integral domain.

(1) A polynomial f in $R[x]$ defines a function $f : R \to R$. If R is infinite, show that f is the zero function (that is, $f(a) = 0$ for all $a \in R$) if and only if f is the zero polynomial.
(2) A polynomial f in $R[x_1, \ldots, x_r]$ defines a function $f : R^r \to R$. If R is infinite, use induction on r to show f is the zero function if and only if f is the zero polynomial.

EXERCISE 3.6.26. Let R be a commutative ring and $S = R[x]$ the polynomial ring in one variable over R. If $W = \{1, x, x^2, \ldots\}$, then the localization S_W is called the *Laurent polynomial ring* over R (see Exercise 3.5.3). Usually, the ring of Laurent polynomials over R is denoted $R[x, x^{-1}]$.

(1) Show that every element of $R[x, x^{-1}]$ has a unique representation in the form $f(x)/x^n$ where $f(x) \in R[x]$ and $n \geq 0$.
(2) If R is an integral domain, prove that the group of units in $R[x, x^{-1}]$ is equal to the set $\{ux^e \mid u \in R^* \text{ and } e \in \mathbb{Z}\}$.
(3) If R is an integral domain, prove that the group of units in $R[x, x^{-1}]$ is the internal direct product $R^* \times \langle x \rangle$.
(4) Let k be a field. Prove that $k[x, x^{-1}]$ is a PID.
(5) Let R be a UFD. Prove that $R[x, x^{-1}]$ is a UFD.

EXERCISE 3.6.27. Let R be a UFD and P a nonzero prime ideal of R. Prove that P contains a prime element π of R. (Hint: Let $x \in P - (0)$. Show that P contains at least one prime divisor of x.)

EXERCISE 3.6.28. (GCD is invariant under a change of base field) Let $k \subseteq F$ be a tower of fields such that k is a subfield of F. In this case we view $k[x]$ as a subring of $F[x]$. Let $f, g \in k[x]$. Prove that if d is the greatest common divisor of f and g in $k[x]$, then d is the greatest common divisor of f and g in $F[x]$.

EXERCISE 3.6.29. Let F be a field of positive characteristic p. Let $\theta : F[y] \to F[y]$ be the evaluation mapping given by $y \mapsto y^p$. Let $F[y^p]$ denote the image of θ. Prove that θ

extends to a homomorphism $\chi : F(y) \to F(y)$ and let $F(y^p)$ be the image of χ. Prove that $F(y^p)$ is the quotient field of $F[y^p]$ and that the diagram

$$\begin{array}{ccc} F[y] & \longrightarrow & F(y) \\ \uparrow & & \uparrow \\ F[y^p] & \longrightarrow & F(y^p) \end{array}$$

commutes where each of the four maps is the set inclusion homomorphism.

EXERCISE 3.6.30. Let $K = F(y^p)$ be the subfield of $L = F(y)$ defined as in Exercise 3.6.29. We say that L/K is an extension of fields. Show that the polynomial $f = x^p - y^p$ is irreducible in $K[x]$, but that $f = (x-y)^p$ in $L[x]$.

EXERCISE 3.6.31. Let p be a prime number and R a commutative ring of characteristic p. Let $R[x,y]$ be the ring of polynomials in two variables with coefficients in R. Prove:

(1) If $n \geq 0$, then $(x+y)^{p^n} = x^{p^n} + y^{p^n}$ in $R[x,y]$. (Hint: Exercise 3.2.30.)
(2) If $n > 0$ and $0 < k < p^n$, then $\binom{p^n}{k}$ is divisible by p.

EXERCISE 3.6.32. Let k be a field. In Algebraic Geometry, the ring $k[x^2,x^3]$ of Exercise 3.6.17 corresponds to a cuspidal cubic curve and is not a UFD. The ring $k[x^2,x+x^3]$ corresponds to a nodal cubic curve.

(1) Show that the quotient field of $k[x^2,x+x^3]$ is $k(x)$. In other words, $k[x^2,x+x^3]$ and $k[x]$ are birational.
(2) Prove that $k[x^2,x+x^3]$ is not a UFD.

EXERCISE 3.6.33. Assume R is a commutative ring and $\theta : R \to A$ is a homomorphism of rings such that the image of θ is a subring of the center of A. Let $a \in A$ and $\sigma : R[x] \to A$ the evaluation map defined by $x \mapsto a$. Let $R[a]$ denote the image of σ. Show that $R[a]$ is the smallest subring of A containing $\theta(R)$ and a. Show that $R[a]$ is commutative.

EXERCISE 3.6.34. Let R be a commutative ring and $a \in R$. Prove that $R[x]/(x-a) \cong R$.

EXERCISE 3.6.35. Let k be an infinite field and assume there exists a monic irreducible polynomial of degree d in $k[x]$. Show that there are infinitely many monic irreducible polynomials of degree d in $k[x]$.

7. Polynomials over a Unique Factorization Domain

PROPOSITION 3.7.1. *(The Rational Root Theorem) Suppose R is a UFD with quotient field K and $u = b/c$ is an element of K such that* $\gcd(b,c) = 1$. *If $f = a_0 + a_1x + \cdots + a_dx^d \in R[x]$ and u is a root of f, then $b \mid a_0$ and $c \mid a_d$.*

PROOF. If $f(b/c) = 0$, then

$$a_0 + \frac{a_1 b}{c} + \frac{a_2 b^2}{c^2} + \cdots + \frac{a_d b^d}{c^d} = 0.$$

Multiply by c^d

$$a_0 c^d + a_1 b c^{d-1} + a_2 b^2 c^{d-2} + \cdots + a_d b^d = 0.$$

Since b divides the last d terms, it follows that $b \mid a_0 c^d$. Since c divides the first d terms, it follows that $c \mid a_d b^d$. Since $\gcd(b,c) = 1$ and R is a UFD, it follows that $b \mid a_0$ and $c \mid a_d$. □

Let R be a unique factorization domain, or UFD for short. Suppose f is a nonzero polynomial in $R[x]$. If we write $f = a_0 + a_1x + \cdots + a_nx^n$, then the *content* of f, written $C(f)$, is defined to be $\gcd(a_0, a_1, \ldots, a_n)$. By Corollary 3.4.13, $C(f)$ is unique up to associates, which means $C(f)$ is unique up to multiplication by a unit of R. If $C(f) = 1$, then we say f is *primitive*.

LEMMA 3.7.2. *Let R be a UFD and f a nonzero polynomial in $R[x]$. If $c_1 = C(f)$, then f factors as $f = c_1f_1$ where $f_1 \in R[x]$ is primitive. The factors c_1 and f_1 of f are unique up to associates in $R[x]$.*

PROOF. By Exercise 3.4.31, if we factor out the content, then $f = C(f)f_1$ where $C(f_1) = 1$. Both $C(f)$ and $C(f_1)$ are unique up to multiplication by units of R. But units of $R[x]$ correspond to the units of R by Exercise 3.6.19. So f_1 is unique up to associates in $R[x]$. □

LEMMA 3.7.3. *Let R be a UFD with quotient field K. Let f and g be nonzero polynomials in $R[x]$.*

(1) If f and g are primitive, then fg is primitive.
(2) $C(fg) = C(f)C(g)$.
(3) Suppose f and g are primitive. Then f and g are associates in $R[x]$ if and only if they are associates in $K[x]$.

PROOF. (1): Assume f and g are nonzero elements of $R[x]$ and fg is not primitive. Then $C(fg)$ is not a unit in R. Let p be an irreducible factor of $C(fg)$ in R. Under the natural map $\eta : R[x] \to R/(p)[x]$ of Theorem 3.6.2 (1), we have $\eta(fg) = \eta(f)\eta(g) = 0$. By Corollary 3.4.14, (p) is a prime ideal, so $R/(p)$ is an integral domain. Thus $R/(p)[x]$ is an integral domain, which implies one of $\eta(f)$ or $\eta(g)$ is zero. That is, p divides the content of f or the content of g. That is, either f or g is not primitive.

(2): As in Lemma 3.7.2, we factor $f = C(f)f_1$, $g = C(g)g_1$, where f_1 and g_1 are primitive. Then $fg = C(f)C(g)f_1g_1$. By Part (1), f_1g_1 is primitive. By Lemma 3.7.2, $C(fg) = C(f)C(g)$.

(3): We are given that $1 = C(f) = C(g)$. Assume f and g are associates in $K[x]$. By Exercise 3.6.19, a unit in $K[x]$ is a nonzero constant polynomial. Suppose $f = ug$ where $u = r/s$ is a unit in K and $\gcd(r, s) = 1$. Then $sf = rg$ implies $sC(f) = rC(g)$, which implies r and s are associates. Therefore u is a unit in R. The converse is trivial, since $R \subseteq K$. □

THEOREM 3.7.4. *(Gauss' Lemma) Let R be a UFD with quotient field K. Suppose $f \in R[x]$ is primitive. Then f is irreducible in $R[x]$ if and only if f is irreducible in $K[x]$.*

PROOF. If f has a nontrivial factorization in $R[x]$, then this factorization still holds in $K[x]$. Assume $f = pq$ is a factorization in $K[x]$, where we assume $m = \deg p \geq 1$, and $n = \deg q \geq 1$. Write

$$p = \sum_{i=0}^{m} \frac{a_i}{b_i}x^i, \quad q = \sum_{i=0}^{n} \frac{c_i}{d_i}x^i$$

and set $b = b_0b_1 \cdots b_m$, $d = d_0d_1 \ldots d_m$. Then $b(a_i/b_i) = \alpha_i \in R$ and $d(c_i/d_i) = \gamma_i \in R$ for each i, so we get

$$bp = \sum_{i=0}^{m} \alpha_i x^i, \quad dq = \sum_{i=0}^{n} \gamma_i x^i$$

are both in $R[x]$. Let $\alpha = C(bp)$ and factor $bp = \alpha p_1$, where p_1 is primitive (Lemma 3.7.2). Set $\gamma = C(dq)$ and factor $dq = \gamma q_1$ where q_1 is primitive (Lemma 3.7.2). Combining all

of this, we have $(bd)f = (\alpha\gamma)(p_1q_1)$. By Lemma 3.7.3, it follows that bd and $\alpha\gamma$ are associates in R. Up to a unit in R, $f = p_1q_1$. □

THEOREM 3.7.5. *Let R be a UFD. Then $R[x_1,\dots,x_n]$ is a UFD.*

PROOF. By finite induction, it is enough to show $R[x]$ is a UFD.

(Existence.) Let $f \in R[x]$ be a nonunit nonzero. If f has degree zero, then we can view f as an element of R and factor f into irreducibles in R. This is a factorization into irreducibles in $R[x]$.

Assume $\deg f \geq 1$ and factor $f = C(f)f_1$ where f_1 is primitive and $C(f) \in R$. Since $C(f)$ can be factored into irreducibles, we can reduce to the case where f is primitive. Let K be the quotient field of R. We know that $K[x]$ is a UFD, by Corollary 3.6.5. Let $f = p_1 \cdots p_n$ be the unique factorization of f into a product of irreducibles in $K[x]$. By Theorem 3.7.4, for each i we can write

$$p_i = \frac{a_i}{b_i} q_i$$

where $a_i, b_i \in R$, and $q_i \in R[x]$ is primitive and irreducible. Set $\alpha = a_1 \cdots a_n$ and $\beta = b_1 \cdots b_n$. Multiplying,

$$f = \frac{\alpha}{\beta} q_1 q_2 \cdots q_n.$$

By Lemma 3.7.3 (3) we conclude that α and β are associates in R. Up to associates, we have factored $f = q_1q_2 \cdots q_n$ into irreducibles in $R[x]$.

(Uniqueness.) Let f be a nonzero nonunit element of $R[x]$. Then f can be factored into a product of irreducibles $f = (c_1 \cdots c_m)(p_1p_2 \cdots p_n)$ where each p_i is a primitive irreducible polynomial in $R[x]$ and each c_i is an irreducible element of R. Up to associates, $C(f) = c_1c_2 \cdots c_m$ is uniquely determined by f. Since R is a UFD, the factorization $C(f) = c_1c_2 \cdots c_m$ is unique in R. In $K[x]$ the factorization $p_1p_2 \cdots p_n$ is uniquely determined up to associates. By Lemma 3.7.3 (3), the factorization is unique in $R[x]$. □

THEOREM 3.7.6. *(Eisenstein's Irreducibility Criterion) Let R be UFD and $f = a_0 + a_1x + \cdots + a_nx^n$ a primitive polynomial of degree $n \geq 1$ in $R[x]$. Let p be a prime in R such that $p \nmid a_n$, $p \mid a_i$ for $i = 0, 1, \dots, a_{n-1}$, and $p^2 \nmid a_0$. Then f is irreducible.*

PROOF. Let $P = (p)$. Then P is a prime ideal in R by Corollary 3.4.14. The proof is by contraposition. Assume $a_n \notin P$, $(a_0,\dots,a_{n-1}) \subseteq P$ and f is reducible. We prove that $p^2 \mid a_0$. By assumption, there is a factorization $f = gh$, where $\deg g = s \geq 1$, $\deg h = t \geq 1$, and $s + t = n$. By Theorem 3.6.2 (1) the natural map $\eta : R \to R/P$ induces $\bar{\eta} : R[x] \to R/P[x]$. Under this homomorphism, $\bar{\eta}(f) = \bar{\eta}(g)\bar{\eta}(h)$. By hypothesis, $\bar{\eta}(f) = \eta(a_n)x^n$ has degree n. If we write $g = b_0 + b_1x + \cdots + b_sx^s$ and $h = c_0 + c_1x + \cdots + c_tx^t$, then

$$\eta(a_n)x^n = (\eta(b_0) + \eta(b_1)x + \cdots + \eta(b_s)x^s)(\eta(c_0) + \eta(c_1)x + \cdots + \eta(c_t)x^t) \tag{7.1}$$

holds in $R/P[x]$. Since P is prime, R/P is an integral domain. Let K denote the quotient field of R/P. The factorization of $\bar{\eta}(f)$ in (7.1) holds in $K[x]$. By Corollary 3.6.5, $K[x]$ is a UFD. We conclude that $(b_0, b_1, \dots, b_{s-1}) \subseteq P$ and $(c_0, c_1, \dots, c_{t-1}) \subseteq P$. In particular, $p \mid b_0$ and $p \mid c_0$. The constant term of f is equal to $a_0 = b_0c_0$ which is divisible by p^2. □

EXAMPLE 3.7.7. Let k be a field and $f(x) \in k[x]$. Assume $\deg f \geq 2$. The set of zeros of $y^2 - f(x)$ in k^2 is called an affine hyperelliptic curve. Assume f is square-free. In other words, f is not divisible by the square of an irreducible polynomial. By Theorem 3.7.6, $y^2 - f(x)$ is irreducible in $k[x,y]$.

EXAMPLE 3.7.8. Let $\Phi(x) = x^p - 1 \in \mathbb{Z}[x]$. Consider $\phi(x) = \Phi(x)/(x-1) = x^{p-1} + x^{p-2} + \cdots + x + 1$. By Exercise 3.6.22, the change of variable $x = y+1$ induces an isomorphism $\mathbb{Z}[x] \cong \mathbb{Z}[y]$. Applying the Binomial Theorem (Exercise 3.1.22) we see that

$$\begin{aligned}\phi(y+1) &= \frac{\Phi(y+1)}{y} \\ &= \frac{(y+1)^p - 1}{y} \\ &= y^{p-1} + \binom{p}{1} y^{p-2} + \cdots + \binom{p}{p-2} y + \binom{p}{p-1}.\end{aligned}$$

By Exercise 1.2.21, p divides $\binom{p}{i}$ if $1 \leq i \leq p-1$. By Theorem 3.7.6, $\phi(y+1)$ is irreducible in $\mathbb{Z}[y]$. Therefore, $\phi(x)$ is irreducible in $\mathbb{Z}[x]$ and by Gauss' Lemma (Theorem 3.7.4), $\phi(x)$ is irreducible in $\mathbb{Q}[x]$.

7.1. Rational Function Fields. Let k be a field and x, y indeterminates. Let $K = k(x)$ be the field of rational functions over k in the variable x. A rational function $\phi \in K$ can be written as a quotient $\phi = p/q$ where $p, q \in k[x]$ are polynomials and $\gcd(p,q) = 1$. By unique factorization in $k[x]$, the polynomials p and q are uniquely determined up to associates. If $u \in k$, and $q(u) \neq 0$, then $\phi(u) = p(u)q(u)^{-1}$ is an element of k. The *pole set* of ϕ is the set of roots of q and the *zero set* of ϕ is the set of roots of p. If u is not a pole of ϕ, then $f(u) = p(u)q(u)^{-1}$ is a well defined element of k. So if the pole set of ϕ is not equal to k, ϕ defines a function on the complement of its pole set. The next theorem provides an Eisenstein irreducibility criterion for polynomials in $K[y]$. It first appeared in [**11**].

THEOREM 3.7.9. *Let k be a field and x, y indeterminates. Let $K = k(x)$ be the field of rational functions over k in the variable x. Let $f(y) = f_0 + f_1 y + f_2 y^2 + \cdots + f_n y^n$ be a polynomial in $K[y]$ where $n \geq 1$ and $f_n \neq 0$. If*

(1) each f_i is a polynomial in $k[x]$,
(2) x divides each of $f_0, f_1, \ldots, f_{n-1}$ and x does not divide f_n, and
(3) x^2 does not divide f_0,

then f is irreducible in $K[y]$.

PROOF. For sake of contradiction, suppose

$$f = (a_0 + a_1 y + \cdots + a_r y^r)(b_0 + b_1 y + \cdots + b_s y^s) \tag{7.2}$$

where $r \geq 1$, $s \geq 1$, and each a_i and b_j is in $K = k(x)$. We have

$$\begin{aligned} f_0 &= a_0 b_0 \\ f_1 &= a_0 b_1 + a_1 b_0 \\ f_2 &= a_0 b_2 + a_1 b_1 + a_2 b_0 \\ &\vdots \\ f_n &= a_0 b_n + \cdots + a_n b_0 \end{aligned}$$

By hypothesis (2), $0 = f_0(0) = f_1(0) = \cdots = f_{n-1}(0)$ and $f_n(0) \neq 0$. We start with $0 = f_0(0) = (a_0 b_0)(0)$. Write $a_0 = p/q$, $b_0 = g/h$, where p, q, g, h are polynomials in $k[x]$ and $\gcd(p,q) = \gcd(g,h) = 1$. Then $pq = f_0 qh$ in $k[x]$. Since $x \mid f_0$ we have $x \mid p$ or $x \mid q$. Suppose for contradiction's sake that $x \mid p$ and $x \mid q$. Then x does not divide q and x does

not divide h. Thus x^2 divides f_0, a contradiction. Assume from now on that $x \mid a_0$ and x does not divide b_0. Equivalently, assume $a_0(0) = 0$ and $b_0(0) \neq 0$. Now we consider

$$0 = f_1(0) = (a_0 b_1)(0) + (a_1 b_0)(0). \tag{7.3}$$

From step one, $a_0(0) = 0$, $b_0(0) \neq 0$, hence (7.3) reduces to $0 = a_1(0)$. Now look at

$$0 = f_2(0) = (a_0 b_2)(0) + (a_1 b_1)(0) + (a_2 b_0)(0) \tag{7.4}$$

which reduces to $a_2(0) = 0$ by applying the first two steps. Iterating this argument, we see that $0 = a_0(0) = a_1(0) = a_2(0) = \cdots = a_r(0)$. This implies $f_n(0) = 0$, a contradiction. □

7.2. Exercises.

EXERCISE 3.7.10. Let $n \in \mathbb{Z}$ and consider the polynomial $f(x) = x^3 + nx - 2$. Show that $f(x)$ is reducible over $\mathbb{Q}$ if and only if n is in the set $\{1, -3, -5\}$.

EXERCISE 3.7.11. Let $f(x) = 20x^5 + 35x^4 - 42x^3 + 21x^2 + 70$ and $g(x) = 80x^5 + 18x^3 - 24x - 15$. Let $F = \mathbb{Q}[x]/(f)$ and $G = \mathbb{Q}[x]/(g)$. Show that F and G are fields.

EXERCISE 3.7.12. Modify the method of Example 3.7.8 to show that the following polynomials are irreducible over $\mathbb{Q}$.

(1) $x^4 + 1$
(2) $x^4 + a^2$, where $a \in \mathbb{Z}$ is odd.
(3) $x^8 + 1$
(4) $x^9 + 2$
(5) $x^{2^n} + a^2$, where $a \in \mathbb{Z}$ is odd and $n \geq 1$.
(6) $x^{p^n} + p - 1$, where p is prime and $n \geq 1$.

(Hint: For (5) and (6), apply Exercise 3.6.31.)

EXERCISE 3.7.13. Let k be a field. If $f(x) = a_0 + a_1 x + \cdots + a_n x^n$ and $a_n \neq 0$, then the *reverse* of f is the polynomial $f^r(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n$.

(1) Show that $f^r(x) = x^n f(x^{-1})$.
(2) If $a_0 \neq 0$, show that f is irreducible over k if and only if f^r is irreducible over k.

EXERCISE 3.7.14. Let $f = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} + a_n x^n$ be a polynomial of degree $n \geq 1$ in $\mathbb{Z}[x]$. Let p be a prime and $[f] = [a_0] + [a_1]x + [a_2]x^2 + \cdots + [a_{n-1}]x^{n-1} + [a_n]x^n$ be the polynomial over the prime field $\mathbb{Z}/(p)$ achieved by reducing the coefficients of f modulo p.

(1) If $[f]$ has degree n and is irreducible over $\mathbb{Z}/(p)$, then f is irreducible over $\mathbb{Q}$. Proof:
(2) Show by counterexample that (a) is false if the degree of $[f]$ is less than n.
(3) Show by counterexample that the converse of (a) is false.

EXERCISE 3.7.15. Let $f = x^3 + 1$. Prove that there is an isomorphism $\theta : \mathbb{Q}[x]/(f) \to F_1 \oplus F_2$ where F_1 and F_2 are fields. Carefully describe the fields F_1 and F_2, and the map θ.

EXERCISE 3.7.16. Let k be a field, a, b, c some elements of k and assume $a \neq b$. Let $f = (x-a)(x-b)$ and $g = (x-c)^2$. Prove:

(1) The ring $k[x]/(x-a)$ is isomorphic to k.
(2) There is an isomorphism of rings $k[x]/(f) \cong k \oplus k$.
(3) There is an isomorphism of rings $k[x]/(g) \cong k[x]/(x^2)$.
(4) If h is a monic irreducible quadratic polynomial in $k[x]$, then the rings $k[x]/(f)$, $k[x]/(g)$, and $k[x]/(h)$ are pairwise nonisomorphic.

EXERCISE 3.7.17. Let k be a field. In this exercise we outline a proof that a rational function in one variable over k has a partial fraction decomposition. Prove:

(1) If f and g are two nonzero polynomials in $k[x]$ and $d = \gcd(f,g)$, then there exist polynomials u, v in $k[x]$ such that $d = fu + gv$, $\deg u < \deg g$, and $\deg v < \deg f$.
(2) If $1 = \gcd(f,g)$ and $\deg h < \deg(fg)$, then there exist unique polynomials u and v satisfying:
$$\frac{h}{fg} = \frac{u}{f} + \frac{v}{g},$$
$\deg u < \deg f$, and $\deg v < \deg g$.
(3) Let g be a polynomial of degree at least one. Let
$$g = p_1^{e_1} \cdots p_n^{e_n}$$
be the unique factorization of g where $p_1, \ldots, p_n$ are distinct irreducibles, $n \geq 2$, and $e_i \geq 1$ for each i. If f is a polynomial and $\deg f < \deg g$, then there exist unique polynomials $q_1, \ldots, q_n$ satisfying:
$$\frac{f}{g} = \frac{q_1}{p_1^{e_1}} + \cdots + \frac{q_n}{p_n^{e_n}},$$
and for each i, $\deg q_i < \deg p_i^{e_i}$.
(4) Let g be a polynomial of degree at least one, $n \geq 1$, and f a polynomial satisfying $\deg f < \deg g^n$. Then there exist unique polynomials $f_0, \ldots, f_{n-1}$ satisfying:
$$f = f_0 + f_1 g + \cdots + f_{n-1} g^{n-1}$$
and for each i, $\deg f_i < \deg g$.
(5) Let g be a polynomial of degree at least one, $n \geq 1$, and f a polynomial satisfying $\deg f < \deg g^n$. Then there exist unique polynomials $f_0, \ldots, f_{n-1}$ satisfying:
$$\frac{f}{g^n} = \frac{f_0}{g^n} + \frac{f_1}{g^{n-1}} + \cdots + \frac{f_{n-1}}{g},$$
and for each i, $\deg f_i < \deg g$.

CHAPTER 4

Linear Algebra

"What makes Linear Algebra linear?" is an important question that every student of this subject should be prepared to answer. I have not forgotten the first time I was asked this question. It was the beginning of the semester when I was taking my first undergraduate course on Linear Algebra. I was living on campus, and at the dining hall one evening one of the people at our table asked the above question. The event has stuck with me because I did not have an answer for my friend. Here is the answer to the question, and the response you should give when you are asked. Algebra is the study of polynomial equations and in this light, Linear Algebra is the study of linear equations.

As much as possible, we study linear algebra over a general ring. Nevertheless, because of the introductory nature of this book, most of the results assume the ground ring is commutative. We hope that a reasonable balance has been achieved between accessibility and generality of results. We define a module over an arbitrary ring and a vector space over a division ring. Algebras are defined over commutative rings. The basis theorem for finitely generated modules is proved for modules over a euclidean domain. The isomorphism between the ring of endomorphisms of a finitely generated free module and the ring of matrices is constructed for an arbitrary commutative ring.

1. Modules

1.1. Definitions and First Properties. In this section we introduce the notion of a module over an arbitrary ring R. An abelian group M is an R-module if multiplication by elements of R turns R into a ring of endomorphisms of M.

DEFINITION 4.1.1. If R is a ring, a *left R-module* is a nonempty set M with an addition operation making M an abelian group together with a left multiplication action by R such that for all $r,s \in R$ and $x,y \in M$ the rules

(1) $r(x+y) = rx+ry$
(2) $r(sx) = (rs)x$
(3) $(r+s)x = rx+sx$
(4) $1x = x$

are satisfied. If R is a field, then M is called a *vector space*.

By default, an R-module is assumed to be a left R-module. This is in agreement with our convention that functions act from the left (Section 1.1.2). There will be times when for sake of convenience we will utilize right R-modules. The statement of the counterpart of Definition 4.1.1 for a right R-module is left to the reader. In Lemma 2.4.1 we saw that a group G acts on a set X if and only if there is a homomorphism of G into $\mathrm{Perm}(X)$. Lemma 4.1.2 is the counterpart of this notion in the context of modules. By Exercise 2.8.11, if M is an abelian group, then the set of all endomorphisms of M, $\mathrm{Hom}(M,M)$, is a ring. Endomorphisms are added point-wise and multiplication is composition of functions.

LEMMA 4.1.2. *Let R be a ring an M an additive abelian group. The following are equivalent.*

(1) M is an R-module.
(2) There is a homomorphism of rings $\theta : R \to \mathrm{Hom}(M,M)$.

PROOF. (2) implies (1): Instead of $\theta(r)(x)$ we will write $r*x$. This defines a left multiplication action by R on M. Then

$$r*(x+y) = \theta(r)(x+y) = \theta(r)(x) + \theta(r)(y)) = r*x + r*y$$

is Part (1) of Definition 4.1.1,

$$r*(s*x) = \theta(r)(\theta(s)(x) = (\theta(r)\theta(s))(x) = \theta(rs)(x)) = (rs)*x$$

is Part (2),

$$(r+s)*x = \theta(r+s)(x) = (\theta(r)+\theta(s))(x) = \theta(r)(x) + \theta(s)(x) = r*x + s*x$$

is Part (3), and lastly,

$$1*x = \theta(1)(x) = 1_M(x) = x$$

is Part (4).

(1) implies (2): For each $r \in R$, define $\lambda_r : M \to M$ to be the "left multiplication by r" function defined by $\lambda_r(x) = rx$. By the first distributive law, $\lambda_r(x+y) = r(x+y) = rx + ry = \lambda_r(x) + \lambda_r(y)$, so $\lambda_r \in \mathrm{Hom}(M,M)$. Define $\theta : R \to \mathrm{Hom}(M,M)$ by $\theta(r) = \lambda_r$. The associative law implies $\lambda_{rs}(x) = (rs)x = r(sx)$, so $\theta(rs) = \theta(r)\theta(s)$ and θ is multiplicative. By the second distributive law, $\lambda_{r+s}(x) = (r+s)x = rx + sx = \lambda_r(x) + \lambda_s(x)$, so $\theta(r+s) = \theta(r) + \theta(s)$ and θ is additive. Lastly, $\lambda_1 = 1_M$, so $\theta(1) = 1$, hence θ is a homomorphism of rings. □

DEFINITION 4.1.3. Let R be a ring, M an R-module, and $\theta : R \to \mathrm{Hom}(M,M)$ the homomorphism of Lemma 4.1.2. The kernel of θ is denoted $\mathrm{annih}_R(M)$ and is called the *annihilator of M in R*. Then $\mathrm{annih}_R(M)$ is equal to $\{r \in R \mid rx = 0 \text{ for all } x \in M\}$. Since θ is a homomorphism of rings, $\mathrm{annih}_R(M)$ is a two-sided ideal in R. If θ is one-to-one, then we say M is a *faithful R-module*.

EXAMPLE 4.1.4. Standard examples of modules are listed here.

(1) If R is any ring, and I is a left ideal in R, then R acts on I from the left. If $x \in I$ and $r \in R$, then $rx \in I$. The associative and distributive laws in R apply. Thus I is an R-module.
(2) Let M be any additive abelian group. Then $\mathbb{Z}$ acts on M. If $x \in M$ and $n \in \mathbb{Z}$, then

$$nx = \begin{cases} 0 & \text{if } n = 0 \\ \sum_{i=1}^{n} x = x + x + \cdots + x & \text{if } n > 0 \\ -\sum_{i=1}^{|n|} x = -(x + x + \cdots + x) & \text{if } n < 0 \end{cases}$$

Using Exercise 2.3.16, the reader should verify that this action makes M into a $\mathbb{Z}$-module.
(3) Let A be an abelian group written additively. Let $m > 1$ be an integer and assume $mx = 0$ for all $x \in A$. It follows from Exercise 4.1.32 that A is a $\mathbb{Z}/m$-module by the action $[n]x = nx$. In particular, if p is a prime and $px = 0$ for all $x \in A$, then A is a vector space over the field $\mathbb{Z}/p$.
(4) Let $\phi : R \to S$ be a homomorphism of rings. Then R acts on S by the multiplication rule $rx = \phi(r)x$, for $r \in R$ and $x \in S$. By this action, S is an R-module.

(5) Let $\phi : R \to S$ be a homomorphism of rings. If M is an S-module, then R acts on M by the multiplication rule $rx = \phi(r)x$, for $r \in R$ and $x \in M$. By this action, M is an R-module.

LEMMA 4.1.5. *Let M be an R-module, $x \in M$, and $r \in R$. Then the following are true:*

(1) $r0 = 0$.
(2) $0x = 0$.
(3) $-1x = -x$.

PROOF. (1): $r0 = r(0+0) = r0 + r0$. Since $M, +$ is a group, we cancel $r0$ to get $r0 = 0$.

(2): $0x = (0+0)x = 0x + 0x$. Since $M, +$ is a group, we cancel $0x$ to get $0x = 0$.

(3): $0 = (1-1)x = 1x + (-1)x = x + (-1)x$. Since $M, +$ is a group, we get $-x = (-1)x$. □

DEFINITION 4.1.6. Let R be a ring and M an R-module. A *submodule* of M is a nonempty subset $N \subseteq M$ such that N is an R-module under the operation by R on M. If $X \subseteq M$, the *submodule of M generated by X* is

$$\left\{ \sum_{i=1}^{n} r_i x_i \mid n \geq 1, r_i \in R, x_i \in X \right\}.$$

The reader should verify that the submodule generated by X is equal to the intersection of the submodules of M containing X. A submodule is *principal*, or *cyclic*, if it is generated by a single element. The submodule generated by X is denoted (X). If $X = \{x_1, x_2, \ldots, x_n\}$ is finite, we sometimes write $(X) = Rx_1 + Rx_2 + \cdots + Rx_n$. We say M is *finitely generated* if there exists a finite subset $\{x_1, \ldots, x_n\} \subseteq M$ such that $M = Rx_1 + \cdots + Rx_n$.

DEFINITION 4.1.7. If I is a left ideal of R and M is an R-module, then IM denotes the R-submodule of M generated by the set $\{rx \mid r \in I, x \in M\}$. Notice that a typical element of IM is not a product rx, but a finite sum of the form $r_1x_1 + \cdots + r_nx_n$.

DEFINITION 4.1.8. Let R be a ring and M an R-module. If A and B are R-submodules of M, then $A + B$ denotes the R-submodule generated by the set $A \cup B$.

DEFINITION 4.1.9. If M and N are R-modules, an *R-module homomorphism* from M to N is a function $f : M \to N$ satisfying

(1) $f(x+y) = f(x) + f(y)$ and
(2) $f(rx) = rf(x)$

for all $x, y \in M$ and $r \in R$. The *kernel* of the homomorphism f is $\ker(f) = \{x \in M \mid f(x) = 0\}$. The *image* of the homomorphism f is $\mathrm{im}(f) = \{f(x) \in N \mid x \in M\}$. An *epimorphism* is a homomorphism that is onto. A *monomorphism* is a homomorphism that is one-to-one. An *isomorphism* is a homomorphism $f : M \to N$ that is one-to-one and onto. In this case we say M and N are *isomorphic*. An *endomorphism* of M is a homomorphism from M to M.

PROPOSITION 4.1.10. *If $f : M \to N$ is an R-module homomorphism, then the following are true:*

(1) The kernel of f is a submodule of M.
(2) f is one-to-one if and only if $\ker(f) = (0)$.
(3) If A is a submodule of M, then $f(A)$, the image of A under f, is a submodule of N.

(4) If B is a submodule of N, then $f^{-1}(B)$, the preimage of B under f, is a submodule of M.

PROOF. Let A be a submodule of M and B a submodule of N. Since f is a homomorphism of additive groups, $\ker(f)$ is a subgroup of $M,+$, $f(A)$ is a subgroup of $N,+$, and $f^{-1}(B)$ is a subgroup of $M,+$, by Exercise2.3.15. Part (2) follows from the corresponding result for group homomorphisms, Lemma 2.3.7. Let $x \in \ker(f)$ and $r \in R$. Then $f(rx) = rf(x) = r0 = 0$ by Lemma 4.1.5. This completes Part (1). If x is an arbitrary element of A, then $f(x)$ represents a typical element of $f(A)$. Then $rf(x) = f(rx) \in f(A)$, which completes Part (3). Let $x \in M$ such that $f(x) \in B$. Then x represents a typical element of $f^{-1}(B)$. Then $f(rx) = rf(x) \in B$, which completes Part (4). □

DEFINITION 4.1.11. Let R be a ring, M an R-module and S a submodule. The *factor* module of M modulo S is the set $M/S = \{a+S \mid a \in M\}$ of all left cosets of S in M. We sometimes call M/S the quotient module of M modulo S. We define addition and scalar multiplication of cosets by the rules

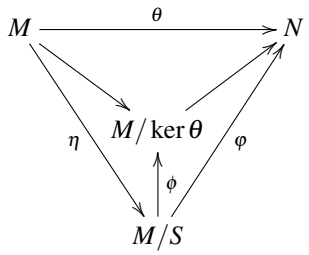
$$(a+S)+(b+S) = (a+b)+S$$
$$r(a+S) = ra+S.$$

The reader should verify that M/S is an R-module. Let $\eta : M \to M/S$ be the natural map defined by $x \mapsto x+S$. Then η is a homomorphism, $\operatorname{im}\eta = M/S$, and $\ker\eta = S$.

Theorem 4.1.12, Corollary 4.1.13, and Theorem 4.1.14 are the counterparts for modules of Theorems 2.3.11, 2.3.12 and 2.3.13.

THEOREM 4.1.12. *(Fundamental Theorem on Homomorphisms of Modules) Let $\theta : M \to N$ be a homomorphism of R-modules. Let S be a submodule of M contained in* $\ker\theta$. *There exists a homomorphism $\varphi : M/S \to N$ satisfying the following.*

(a) $\varphi(a+S) = \theta(a)$, or in other words $\theta = \varphi\eta$.
(b) φ is the unique homomorphism from $M/S \to N$ such that $\theta = \varphi\eta$.
(c) $\operatorname{im}\theta = \operatorname{im}\varphi$.
(d) $\ker\varphi = \eta(\ker\theta) = \ker(\theta)/S$.
(e) φ is one-to-one if and only if $S = \ker\theta$.
(f) φ is onto if and only if θ is onto.
(g) There is a unique homomorphism $\phi : M/S \to M/\ker\theta$ such that the diagram



commutes.

PROOF. On the additive groups, this follows straight from the Fundamental Theorem on Group Homomorphisms, Theorem 2.3.11. The rest is left to the reader. □

COROLLARY 4.1.13. *(The Isomorphism Theorems) Let M be an R-module with submodules A and B.*

(a) The natural map

$$\frac{A}{A\cap B}\to\frac{A+B}{B}$$

sending the coset $x+A\cap B$ to the coset $x+B$ is an isomorphism.

(b) If $A\subseteq B$, then B/A is a submodule of M/A and the natural map

$$\frac{M/A}{B/A}\to M/B$$

sending the coset containing $x+A$ to the coset $x+B$ is an isomorphism.

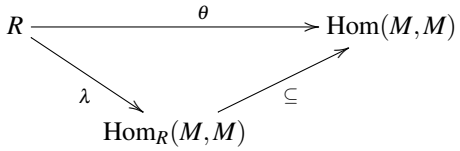
PROOF. This follows from Theorem 4.1.12 and Theorem 2.3.12, its counterpart for groups. □

THEOREM 4.1.14. *(The Correspondence Theorem) Let M be an R-module and A a submodule of M. There is a one-to-one order-preserving correspondence between the submodules B such that $A\subseteq B\subseteq M$ and the submodules of M/A given by $B\mapsto B/A$.*

PROOF. This follows from Proposition 4.1.10 and The Correspondence Theorem for Groups, Theorem 2.3.13. □

DEFINITION 4.1.15. If M and N are R-modules, the set of all R-module homomorphisms from M to N is denoted $\mathrm{Hom}_R(M,N)$. Modules are additive abelian groups and an abelian group has a natural structure as a $\mathbb{Z}$-module (Example 4.1.4 (2)). The set of all group homomorphisms from M to N is denoted $\mathrm{Hom}(M,N)$ or $\mathrm{Hom}_{\mathbb{Z}}(M,N)$. By Exercise 2.8.11, $\mathrm{Hom}_{\mathbb{Z}}(M,N)$ is an abelian group where addition of functions is defined point-wise. Since an R-module homomorphism $\phi:M\to N$ is a homomorphism of additive abelian groups, there is a set containment $\mathrm{Hom}_R(M,N)\subseteq\mathrm{Hom}_{\mathbb{Z}}(M,N)$. Hence $\mathrm{Hom}_R(M,N)$ is an abelian group. The reader should be advised that when R is noncommutative, $\mathrm{Hom}_R(M,N)$ is not an R-module per se. If $M=N$, then in Exercise 4.1.33 the reader is asked to prove that $\mathrm{Hom}_R(M,M)$ is a ring. In general, $\mathrm{Hom}_R(M,M)$ is a noncommutative ring.

EXAMPLE 4.1.16. Let R be a commutative ring and M an R-module. If $r\in R$, then "left multiplication by r" is the function $\lambda_r:M\to M$, where $\lambda_r(x)=rx$. As in Lemma 4.1.2, there is a homomorphism of rings $\theta:R\to\mathrm{Hom}(M,M)$ defined by $\theta(r)=\lambda_r$. Since R is commutative, if $r,s\in R$, then $\lambda_r(sx)=r(sx)=(rs)x=(sr)x=s(rx)=s\lambda_r(x)$. Therefore, λ_r is an R-module homomorphism from M to M. This shows that the homomorphism θ factors through a homomorphism $\lambda:R\to\mathrm{Hom}_R(M,M)$ which we call the *left regular representation* of R in $\mathrm{Hom}_R(M,M)$. The diagram of ring homomorphisms



commutes. For any $\phi\in\mathrm{Hom}_R(M,M)$, $r\in R$, and $x\in M$, $r\phi(x)=\phi(rx)$. Therefore, $\lambda_r\phi=\phi\lambda_r$, which implies the image of R under the homomorphism λ is a subring of the center of $\mathrm{Hom}_R(M,M)$. By θ, $\mathrm{Hom}_R(M,M)$ is turned into an R-algebra (see Definition 4.4.1).

1.2. Direct Sums of Modules. We limit our attention to direct products and direct sums over a finite index set.

DEFINITION 4.1.17. Let R be a ring and $M_1,\dots,M_n$ a finite set of R-modules. The *direct product* of $M_1,\dots,M_n$ is the R-module with underlying set $M_1\times\cdots\times M_n$ and with addition and R-action defined coordinate-wise:

$$(x_1,\dots,x_n)+(y_1,\dots,y_n)=(x_1+y_1,\dots,x_n+y_n)$$
$$r(x_1,\dots,x_n)=(rx_1,\dots,rx_n).$$

The direct product of a finite set of R-modules is frequently called the *(external) direct sum* and is denoted $M_1\oplus M_2\oplus\cdots\oplus M_n$. As in Definition 2.5.1, for each k there are the canonical injection and projection maps

$$M_k \xrightarrow{\iota_k} M_1\oplus M_2\oplus\cdots\oplus M_n \xrightarrow{\pi_k} M_k$$

such that $\pi_k\iota_k = 1_{M_k}$.

DEFINITION 4.1.18. Let $\{S_1,\dots,S_n\}$ be a set of submodules in the R-module M. The submodule of M generated by the set $S_1\cup S_2\cup\cdots\cup S_n$ is called the *sum* of the submodules and is denoted $S_1+S_2+\cdots+S_n$. We say that M is the *internal direct sum* of the submodules in case

(1) $M = S_1+S_2+\cdots+S_n$, and
(2) if $x_i\in S_i$ for each i and $0 = x_1+x_2+\cdots+x_n$, then $x_i = 0$ for each i.

We denote the internal direct sum by $M = S_1\oplus S_2\oplus\cdots\oplus S_n$. If M is an R-module and N is an R-submodule of M, then N is a *direct summand* of M if there is a submodule L of M such that $M = N\oplus L$.

LEMMA 4.1.19. *If $S_1,\dots,S_n$ are submodules in the R-module M, and $M = S_1\oplus\cdots\oplus S_n$, then the following are true.*

(1) For each k, $S_k\bigcap\left(\sum_{j\neq k}S_j\right) = (0)$.
(2) M is isomorphic to the (external) direct sum $S_1\oplus\cdots\oplus S_n$.

PROOF. The proof of the counterpart of this lemma for groups applies here (see Lemma 2.5.4). □

PROPOSITION 4.1.20. *Suppose $S_1,\dots,S_n$ are submodules in the R-module M satisfying*

(1) $M = S_1+S_2+\cdots+S_n$ and
(2) for each k, $S_k\cap\left(\sum_{j\neq k}S_j\right) = (0)$.

Then $M = S_1\oplus S_2\oplus\cdots\oplus S_n$.

PROOF. The proof of the counterpart of this lemma for groups, Proposition 2.5.5, applies here. □

PROPOSITION 4.1.21. *Let R be a ring, M an R-module, and N an R-submodule of M. The following are equivalent.*

(1) N is a direct summand of M.
(2) There is an R-module homomorphism $\pi : M\to N$ such that $\pi(x) = x$ for every $x\in N$.

PROOF. (1) implies (2): There is a submodule L of M such that $M = N\oplus L$. The canonical projection map $\pi : M\to N$ is an R-module homomorphism and $\pi(x) = x$ for every $x\in N$.

(2) implies (1): Let $L = \ker\pi$. Given $z \in M$, let $x = \pi(z)$ and $y = z - x$. Then $\pi(y) = \pi(z) - \pi(x) = x - x = 0$ implies $y \in L$. This shows $M = N + L$. Let $z \in N \cap L$. Then $z \in L$ implies $\pi(z) = 0$ and $z \in N$ implies $\pi(z) = z$. This shows $N \cap L = (0)$. By Proposition 4.1.20, $M = N \oplus L$. □

1.3. Free modules.

DEFINITION 4.1.22. Let R be any ring. As defined in Definition 4.1.6, an R-module M is finitely generated if there exist elements $x_1, \ldots, x_n$ in M such that for each $m \in M$ there exist $r_1, \ldots, r_n$ in R such that $m = r_1x_1 + \cdots + r_nx_n$. Equivalently, M is finitely generated if there is a finite subset $\{x_1, \ldots, x_n\}$ of M such that $M = Rx_1 + \cdots + R_nx_n$. Thus, M is finitely generated if and only if M is equal to the sum of a finite number of cyclic submodules. If M has a finite generating set, then by the Well Ordering Principle, there exists a generating set with minimal cardinality. We call such a generating set a *minimal generating set*. The *rank of M*, written $\mathrm{Rank}(M)$, is defined to be the number of elements in a minimal generating set.

EXAMPLE 4.1.23. If k is a field and V is a finite dimensional k-vector space, then we will see in Theorem 4.2.4 below that the rank of V as defined in Definition 4.1.22 is equal to $\dim_k(V)$, the dimension of V over k.

DEFINITION 4.1.24. Let R be any ring. By Example 4.1.4 (1), R is a left R-module. If $n \geq 1$, we will write $R^{(n)}$ for the direct sum $R \oplus \cdots \oplus R$ of n copies of R. An R module M is said to be *free of finite rank n* if M is isomorphic to $R^{(n)}$ for some n. In particular, $\mathbb{Z}^{(n)}$ is a free $\mathbb{Z}$-module of rank n. If $X = \{x_1, \ldots, x_n\}$ is a finite subset of M, define $\Sigma_X : R^{(n)} \to M$ by $\Sigma_X(r_1, \ldots, r_n) = r_1x_1 + \ldots r_nx_n$. The reader should verify that Σ_X is an R-module homomorphism. We say X is a *linearly independent set* in case Σ_X is one-to-one. Otherwise, X is said to be a *linearly dependent set*. Let $e_i \in R^{(n)}$ be the n-tuple with 1 in coordinate i and 0 elsewhere. The set $\{e_i \mid 1 \leq i \leq n\}$ is a linearly independent generating set for $R^{(n)}$, and is called the *standard basis for $R^{(n)}$*.

LEMMA 4.1.25. *Let R be any ring and M a nonzero finitely generated R-module. Then M is free if and only if there exists a subset $X = \{b_1, \ldots, b_n\} \subseteq M$ which is a linearly independent generating set for M.*

PROOF. Given a finite linearly independent spanning set $X = \{b_1, \ldots, b_n\}$, define $\Sigma_X : R^{(n)} \to M$ by $\Sigma_X(r_1, \ldots, r_n) = \sum_{i=1}^n r_ib_i$. Because X generates M and is linearly independent, Σ is one-to-one and onto. The converse is left to the reader. □

LEMMA 4.1.26. *Let R be any ring, M a nonzero R-module, and $X = \{x_1, \ldots, x_n\}$ a linearly independent subset of M. Every v in the span of X has a unique representation as a linear combination of the form $\alpha_1x_1 + \cdots + \alpha_nx_n$ where $\alpha_1, \ldots, \alpha_n$ are elements of R.*

PROOF. By Definitions 4.1.6 and 4.1.24, the submodule generated by X is equal to the image of $\Sigma_X : R^{(n)} \to M$. The uniqueness claim is equivalent to the fact that Σ_X is one-to-one. □

DEFINITION 4.1.27. Let R be any ring, M an R-module, and X a subset of M. We say X is a *linearly independent set* in case every finite subset of X is linearly independent. When X is a finite set, this definition agrees with Definition 4.1.24. If X is a linearly independent generating set for M, then we say X is a *free basis* for M. By Lemma 4.1.25, if M is finitely generated, then M has a free basis if and only if M is free. In case M is not necessarily finitely generated, and M has a free basis X, we say M is *free*.

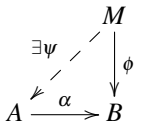
EXAMPLE 4.1.28. We have already seen examples of free modules. Let R be a commutative ring.

(1) The ring of polynomials $R[x]$ is a free R-module and the set $\{1, x, x^2, \ldots, x^i, \ldots\}$ is a free basis.
(2) If G is a group, and $R(G)$ the group ring (see Example 3.1.6), then $R(G)$ is a free R module with free basis $\{g \mid g \in G\}$.

1.4. Projective modules.

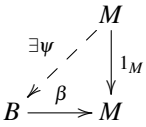
PROPOSITION 4.1.29. *Let R be a ring and M an R-module. The following are equivalent.*

(1) There is a free R-module of finite rank and M is a direct summand of F.
(2) M is finitely generated and for every epimorphism $\beta : B \to M$ of R-modules there exists an R-module homomorphism $\psi : M \to B$ such that $\beta\psi = 1_M$.
(3) M is finitely generated and for any diagram of R-module homomorphisms



with α onto, there exists an R-module homomorphism $\psi : M \to A$ such that $\alpha\psi = \phi$.

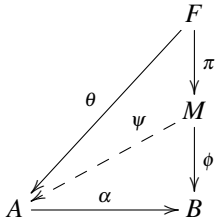
PROOF. (3) implies (2): Consider the diagram



of R-module homomorphisms where $1_M : M \to M$ is the identity map. By (3) there exists $\psi : M \to B$ such that $\beta\psi = 1_M$.

(2) implies (1): Let $X = \{x_1, \ldots, x_n\}$ be a generating set for M. Let $F = R^{(n)}$ be the free R-module of rank n and $\Sigma_X : F \to M$ the R-module epimorphism defined in Definition 4.1.24. By (2) there exists an R-module homomorphism $\psi : M \to F$ such that $\Sigma_X\psi = 1_M$. By Proposition 4.1.21, M is a direct summand of F.

(1) implies (3): Let $F = R^{(n)}$ be a free R-module of rank n and assume M is a direct summand of F. By Proposition 4.1.21, there is an R-module homomorphism $\pi : F \to M$ such that $\pi(x) = x$ for all $x \in M$. Suppose $\phi : M \to B$ and $\alpha : A \to B$ are R-module homomorphisms and α is onto. Let $X = \{x_1, \ldots, x_n\}$ be a basis for F and set $Y = \{y_i = \phi(x_i) \mid 1 \leq i \leq n\}$. Since α is onto, pick $Z = \{z_1, \ldots, z_n\} \subseteq A$ such that $\alpha(z_i) = y_i$. By Exercise 4.1.39 there is a unique R-module homomorphism $\theta : F \to A$ such that $\theta(x_i) = z_i$. Since $\phi\pi(x_i) = y_i = \alpha\theta(x_i)$ and $X = \{x_1, \ldots, x_n\}$ is a generating set for F, we have $\alpha\theta(x) = \phi\pi(x)$ for all $x \in F$. The outer triangle in the diagram



commutes. Define $\psi : M \to A$ to be the restriction of θ to M. If $x \in M$, then $\pi(x) = x$, so $\alpha\psi(x) = \phi(x)$. □

DEFINITION 4.1.30. If R is a ring and M is an R-module satisfying any of the equivalent conditions of Proposition 4.1.29, then we say M is a *finitely generated projective R-module*.

EXAMPLE 4.1.31. Here are some examples of modules that are projective and modules that are not projective.

(1) A free R-module of finite rank satisfies Proposition 4.1.29 (1), hence a finitely generated free R-module is a projective R-module. In particular, R is a free R-module of rank 1.
(2) Let R be a ring containing proper two-sided ideals I and J such that $R = I \oplus J$. Then I and J are direct summands of the free R-module R, hence are projective R-modules by Proposition 4.1.29 (1). By Theorem 3.3.6, $I = Re_1$ and $J = Re_2$, where e_1, e_2 is a set of orthogonal idempotents. Then $e_1e_2 = 0$ is a nontrivial dependence relation. This implies $0 \in J$ does not have a unique representation in terms of any generating set for J. Hence I and J are not free R-modules.
(3) Let p and q be distinct prime numbers. By the Chinese Remainder Theorem, Theorem 1.2.11, $\mathbb{Z}/(pq) \cong \mathbb{Z}/(p) \oplus \mathbb{Z}/(q)$. By Part (2), $\mathbb{Z}/(p)$ is a projective $\mathbb{Z}/(pq)$-module which is not a free $\mathbb{Z}/(pq)$-module.

1.5. Exercises.

EXERCISE 4.1.32. Let R be a commutative ring, I an ideal of R, and M an R-module. As in Definition 4.1.7, IM denotes the R-submodule of M generated by the set $\{rx \mid r \in I, x \in M\}$. Prove that M/IM is an R/I-module under the action $(r+I)(x+IM) = rx+IM$.

EXERCISE 4.1.33. This exercise is based on Exercise 2.8.11. Let M be an R-module, where R is any ring. Follow the outline below to show that the set $\mathrm{Hom}_R(M,M)$ of all R-module endomorphisms of M is a ring.

(1) If $f, g \in \mathrm{Hom}_R(M,M)$, then $f+g$ is the function defined by the rule: $(f+g)(x) = f(x)+g(x)$. Show that this additive binary operation makes $\mathrm{Hom}_R(M,M)$ into an abelian group.
(2) Show that composition of functions defines a binary operation on $\mathrm{Hom}_R(M,M)$ satisfying the following.
   (a) $f(gh) = (fg)h$ for all f, g, h in $\mathrm{Hom}_R(M,M)$. In other words, composition of functions is associative.
   (b) $f(g+h) = fg+fh$ and $(f+g)h = fh+gh$ for all f, g, h in $\mathrm{Hom}_R(M,M)$. In other words, composition distributes over addition.

Together with the two binary operations of addition and composition of endomorphisms, we call $\mathrm{Hom}_R(M,M)$ the *ring of endomorphisms of M*.

EXERCISE 4.1.34. This exercise is based on Exercise 4.1.33. Let M be an R-module, where R is any ring. Let $S = \mathrm{Hom}_R(M,M)$ be the ring of R-module endomorphisms of M. Show that M is a left S-module under the action $\phi x = \phi(x)$, for all $\phi \in S$ and $x \in M$.

EXERCISE 4.1.35. Let R be a commutative ring and I an ideal in R. The natural ring homomorphism $\eta : R \to R/I$ turns R/I into an R-module (Example 4.1.4). Define

$$\phi : \mathrm{Hom}_R(R/I, R/I) \to R/I$$

by $\phi(f) = f(1+I)$. Show that ϕ is an isomorphism of rings.

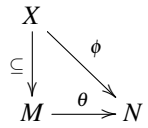
EXERCISE 4.1.36. Let R be a ring and M an R-module. Then M is said to be *simple* if its only submodules are (0) and M.

(1) Prove that any simple R-module is cyclic.
(2) Let M be a non-zero simple R-module. Prove that any R-module homomorphism $h : M \to M$ is either an automorphism of M, or $h(m) = 0$ for every $m \in M$.
(3) (Schur's Lemma) Let M be a non-zero simple R-module. Prove that $\mathrm{Hom}_R(M,M)$ is a division ring.
(4) Say $R = F$ is a field, $M = V$ is a finite dimensional F-vector space. Find necessary and sufficient conditions for V to be simple. Calculate $\mathrm{Hom}_F(V,V)$ for a non-zero simple F-vector space V.
(5) Say $R = \mathbb{Z}$ and M is a finitely generated $\mathbb{Z}$-module. Find necessary and sufficient conditions for M to be simple. Calculate $\mathrm{Hom}_{\mathbb{Z}}(M,M)$ for a non-zero simple $\mathbb{Z}$-module M. (Hint: Corollary 2.2.19.)

EXERCISE 4.1.37. Let R be a ring. The opposite ring of R is defined in Definition 3.1.8. Show that there exists an isomorphism of rings $\mathrm{Hom}_R(R,R) \cong R^o$, where R is viewed as a left R-module and R^o denotes the opposite ring.

EXERCISE 4.1.38. (Module version of Finitely Generated over Finitely Generated is Finitely Generated) Let $R \to S$ be a homomorphism of rings such that S is finitely generated as an R-module. If M is a finitely generated S-module, prove that M is finitely generated as an R-module.

EXERCISE 4.1.39. (Universal Mapping Property) The purpose of this exercise is to prove that a homomorphism on a finitely generated free module is completely determined by its values on a basis. Let R be any ring. Let M and N be R-modules. Assume M is a free R-module of rank m with basis $X = \{x_1, \ldots, x_m\}$. Let $\phi : X \to N$ be any function. Show that there exists a unique homomorphism $\theta \in \mathrm{Hom}_R(M,N)$ such that the diagram



commutes.

EXERCISE 4.1.40. Let F be a field and $R = M_2(F)$ the ring of two-by-two matrices over F. Let

$$e_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Follow the following outline to prove that the ideals Re_1 and Re_2 are finitely generated projective R-modules but not free R-modules.

(1) Show $e_1^2 = e_1$, $e_2^2 = e_2$, $e_1 e_2 = e_2 e_1 = 0$. We say e_1 and e_2 are orthogonal idempotents.
(2) Show that Re_1 is the set of all matrices with second column consisting of zeros.
(3) Show that Re_2 is the set of all matrices with first column consisting of zeros.
(4) Show that $R = Re_1 \oplus Re_2$ as R-modules. Show that Re_i is a finitely generated projective R-module for $i = 1,2$.
(5) For $i = 1,2$, show that Re_i is not a free R-module.

EXERCISE 4.1.41. Let R be any ring and M a free R-module of rank n with basis $X = \{x_1, \ldots, x_n\}$. Use Exercise 4.1.39 to show that the group of units in the ring $\mathrm{Hom}_R(M,M)$ contains a subgroup isomorphic to S_n, the symmetric group on n letters.

EXERCISE 4.1.42. State and prove a version of Exercise 4.1.39 for a free R-module M that is not necessarily finitely generated (see Definition 4.1.27).

2. Vector Spaces

A vector space is a module over a division ring. A submodule of a vector space is called a *subspace*. Elements of a vector space are called a *vectors*. If D is a division ring and V, W are D-vector spaces, then a homomorphism $\phi \in \mathrm{Hom}_D(V,W)$ is called a *linear transformation*. A generating set for V as a D-module is called a *spanning set*.

LEMMA 4.2.1. *Let V be a vector space over a division ring D. If v is a nonzero vector in V, then $\{v\}$ is a linearly independent set. Equivalently, if $v \in V-(0)$, $\alpha \in D$ and $\alpha v = 0$, then $\alpha = 0$.*

PROOF. Assume $\alpha v = 0$ and $\alpha \neq 0$. By Lemma 4.1.5, we have $0 = \alpha^{-1}0 = \alpha^{-1}\alpha v = 1v = v$. □

LEMMA 4.2.2. *Let D be a division ring and V a nonzero finitely generated vector space over D. If $B \subseteq V$, then the following are equivalent.*

(1) B is a basis for V. That is, B is a linearly independent spanning set for V.
(2) B is a spanning set for V and no proper subset of B is a spanning set for V.

PROOF. (1) implies (2): For sake of contradiction, suppose there is a proper subset $B_1 \subsetneq B$ and B_1 is also a spanning set for V. Let $v \in B - B_1$. Since B_1 is a spanning set, there exist $x_1,\ldots,x_n$ in B_1 and $\alpha_1,\ldots,\alpha_n$ in D such that $v = \alpha_1 x_1 + \cdots + \alpha_n x_n$. Then $v - \alpha_1 x_1 - \cdots - \alpha_n x_n = 0$ is a dependency relation in B, which is a contradiction.

(2) implies (1): Assume $B = \{x_1,\ldots,x_n\}$ is a spanning set. We prove that if B is linearly dependent, then there is a proper subset of B that is a spanning set. Since V is nonzero and B is a spanning set, we know B is nonempty. If $0 \in B$, then the span of B is equal to the span of $B - \{0\}$. From now on we assume each x_i is nonzero. Assume $\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$ where $(\alpha_1,\ldots,\alpha_n)$ is a nonzero vector in $D^{(n)}$. Let k be the largest integer satisfying: $\alpha_k \neq 0$ and if $i > k$, then $\alpha_i = 0$. By Lemma 4.2.1, $k > 1$. Then

$$x_k = -\alpha_k^{-1}(\alpha_1 x_1 + \cdots + \alpha_{k-1}x_{k-1})$$

is in the subspace spanned by $x_1,\ldots,x_{k-1}$. Therefore, $B - x_k$ is a spanning set for V. □

COROLLARY 4.2.3. *If V is a finitely generated vector space over a division ring D, then V has a basis.*

PROOF. As in Definition 4.1.22, a minimal generating set exists. By Lemma 4.2.2, a minimal generating set is a basis. □

THEOREM 4.2.4. *Let V be a finitely generated vector space over the division ring D and $B = \{b_1,\ldots,b_n\}$ a basis for V.*

(1) If $Y = \{y_1,\ldots,y_m\}$ is a linearly independent set in V, then $m \leq n$. We can re-order the elements of B such that $\{y_1,\ldots,y_m,b_{m+1},\ldots,b_n\}$ is a basis for V.
(2) Every basis for V has n elements.

PROOF. Step 1: Write $y_1 = \alpha_1 b_1 + \cdots + \alpha_n b_n$ where each $\alpha_i \in D$. For some i, $\alpha_i \neq 0$. Re-order the basis elements and assume $\alpha_1 \neq 0$. Solve for b_1 to get $b_1 = \alpha_1^{-1}y_1 -$

$\sum_{i=2}^{n} \alpha_1^{-1}\alpha_i b_i$. Therefore $B \subseteq Dy_1 + Db_2 + \cdots + Db_n$, hence $\{y_1, b_2, \ldots, b_n\}$ is a spanning set for V. Suppose $0 = \beta_1 y_1 + \beta_2 b_2 + \cdots + \beta_n b_n$. Then

$$\begin{aligned} 0 &= \beta_1 (\alpha_1 b_1 + \cdots + \alpha_n b_n) + \beta_2 b_2 + \cdots + \beta_n b_n \\ &= \beta_1\alpha_1 b_1 + (\beta_1\alpha_2 + \beta_2) b_2 + \cdots + (\beta_1\alpha_n + \beta_n) b_n, \end{aligned}$$

from which it follows that $\beta_1\alpha_1 = 0$, hence $\beta_1 = 0$. Now $0 = \beta_2 b_2 + \cdots + \beta_n b_n$ implies $0 = \beta_2 = \cdots = \beta_n$. We have shown that $\{y_1, b_2, \ldots, b_n\}$ is a basis for V.

Step j: Inductively, assume $j \geq 2$ and that $\{y_1, y_2, \ldots, y_{j-1}, b_j, \ldots, b_n\}$ is a basis for V. Write $y_j = \alpha_1 y_1 + \cdots + \alpha_{j-1} y_{j-1} + \alpha_j b_j + \cdots + \alpha_n b_n$ where each $\alpha_i \in D$. Since the set $\{y_1, \ldots, y_j\}$ is linearly independent, for some $i \geq j$, $\alpha_i \neq 0$. Re-order the basis elements and assume $\alpha_j \neq 0$. Solve for b_j and by a procedure similar to that used in Step 1, we see that $\{y_1, \ldots, y_j, b_{j+1}, \ldots, b_n\}$ is a basis for V.

By finite induction, Part (1) is proved. For Part (2), assume $\{c_1, \ldots, c_m\}$ is another basis for V. By applying Part (1) from both directions, it follows that $m \leq n$ and $n \leq m$. □

DEFINITION 4.2.5. Suppose D is a division ring and V is a vector space over D. If V is finitely generated and nonzero, then we define the *dimension* of V, written $\dim_D(V)$, to be the number of elements in a basis for V. If $V = (0)$, set $\dim_D(V) = 0$ and if V is not finitely generated, set $\dim_D(V) = \infty$.

COROLLARY 4.2.6. *Let V be a finitely generated vector space over the division ring D and $X = \{x_1, \ldots, x_n\}$ a spanning set for V. Then the following are true:*

(1) There is a subset of X that is a basis for V.
(2) $\dim_D V \leq n$.

PROOF. Assume V is nonzero. Then X contains a nonzero vector. Without loss of generality assume $x_1 \neq 0$. By Lemma 4.2.1, $\{x_1\}$ is a linearly independent set. Let S be the set of all subsets of X that are linearly independent. Choose $B \in S$ such that B has maximal cardinality. We show B is a spanning set for V. For sake of contradiction, assume $(B) \neq V$. Since X is a spanning set for V, this implies X is not a subset of (B). Assume $x_n \notin (B)$. Then x_n is not a linear combination of the vectors in B. Therefore, $B \cup \{x_n\}$ is a linearly independent set, which contradicts the maximality of B. □

DEFINITION 4.2.7. Let R be a commutative ring and M a free R-module with a finite basis $\{b_1, \ldots, b_n\}$. By Exercise 4.2.17, any other basis of M has n elements. We call n the *rank* of M and write $\text{Rank}_R M = n$.

PROPOSITION 4.2.8. *(Free over Free is Free) Let $\theta : R \to S$ be a homomorphism of rings such that S is a finitely generated free R-module. Let M be a finitely generated free S-module. As in Example 4.1.4 (4), we view M as an R-module. In this context, M is a finitely generated free R module. If R and S are both commutative, then* $\text{Rank}_R(M) = \text{Rank}_S(M)\text{Rank}_R(S)$.

PROOF. Let $X = \{s_1, \ldots, s_m\}$ be a basis for S over R and $Y = \{y_1, \ldots, y_n\}$ a basis for M over S. Let $Z = \{s_i y_j \mid i = 1, \ldots, m \text{ and } j = 1, \ldots, n\}$. We show Z is basis for M over R.

Step 1: Z is a spanning set for M as an R-module. Let x be an arbitrary element of M. There exist $b_1, \ldots, b_n$ in S such that $x = \sum_{j=1}^{n} b_j y_j$. For each j there exist $a_{1j}, \ldots, a_{mj}$ in R such that $b_j = \sum_{i=1}^{m} a_{ij} s_i$. Taken together, we have

$$x = \sum_{j=1}^{n} b_j y_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} s_i \right) y_j = \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij} (s_i y_j)$$

which shows Z is a spanning set for M over R.

Step 2: Z is linearly independent over R. Assume there is a dependence relation $0 = \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij}(s_i y_j)$ where the elements a_{ij} are in R. Since Y is a basis for M over S, for each j we have $\sum_{i=1}^{m} a_{ij} s_i = 0$ in S. Since X is a basis for S over R, we have $a_{ij} = 0$ for every i and for every j.

The cardinality of Z is equal to $|Z| = |X||Y|$, which proves the rank formula. □

2.1. Exercises.

EXERCISE 4.2.9. Suppose D is a division ring, V is a finite dimensional vector space over D, and W is a subspace of V. Prove:

(1) W is finite dimensional and $\dim_D(W) \leq \dim_D(V)$.
(2) There is a subspace U of V such that $V = U \oplus W$ is an internal direct sum and $\dim_D(V) = \dim_D(U) + \dim_D(W)$.
(3) $\dim_D(V/W) = \dim_D(V) - \dim_D(W)$.

EXERCISE 4.2.10. Suppose $\phi \in \mathrm{Hom}_D(V,W)$, where V and W are vector spaces over the division ring D. Prove:

(1) If V is finite dimensional, then the kernel of ϕ is finite dimensional and the image of ϕ is finite dimensional.
(2) If the kernel of ϕ is finite dimensional and the image of ϕ is finite dimensional, then V is finite dimensional.

EXERCISE 4.2.11. (The Rank-Nullity Theorem) Suppose $\phi \in \mathrm{Hom}_k(V,W)$, where V and W are vector spaces over the field k. The *rank* of ϕ, written $\mathrm{Rank}(\phi)$, is defined to be the dimension of the image of ϕ. The *nullity* of ϕ, written $\mathrm{Nullity}(\phi)$, is defined to be the dimension of the kernel of ϕ. Prove that if V is finite dimensional, then $\dim_k(V) = \mathrm{Rank}(\phi) + \mathrm{Nullity}(\phi)$.

EXERCISE 4.2.12. Suppose $\phi \in \mathrm{Hom}_D(V,V)$, where V is a finite dimensional vector space over the division ring D. Prove that the following are equivalent:

(1) ϕ is invertible.
(2) $\mathrm{Nullity}(\phi) = 0$.
(3) $\mathrm{Rank}(\phi) = \dim_D(V)$.

EXERCISE 4.2.13. Let R be a UFD with quotient field K. Let a be an element of R which is not a square in R and let $f = x^2 - a \in R[x]$.

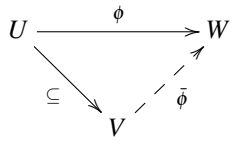
(1) Show that $S = R[x]/(f)$ is an integral domain and $L = K[x]/(f)$ is a field.
(2) Show that S is a free R-module, $\mathrm{Rank}_R(S) = 2$, and $\dim_K(L) = 2$.

EXERCISE 4.2.14. Let V be a finite dimensional vector space over a division ring D. Let ϕ, ψ be elements of $\mathrm{Hom}_D(V,V)$. Prove:

(1) $\mathrm{Rank}(\phi\psi) \leq \mathrm{Rank}(\phi)$.
(2) $\mathrm{Rank}(\phi\psi) \leq \mathrm{Rank}(\psi)$.
(3) $\mathrm{Rank}(\phi\psi) \leq \min(\mathrm{Rank}(\phi), \mathrm{Rank}(\psi))$.
(4) If ϕ is invertible, $\mathrm{Rank}(\phi\psi) = \mathrm{Rank}(\psi\phi) = \mathrm{Rank}(\psi)$.

EXERCISE 4.2.15. Let D be a division ring and V and W finitely generated vector spaces over D. Suppose U is a subspace of V and $\phi : U \to W$ an element of $\mathrm{Hom}_D(U,W)$.

Show that there exists an element $\bar{\phi}$ of $\mathrm{Hom}_D(V,W)$ such that the diagram



commutes. That is, $\bar{\phi}$ is an extension of ϕ.

EXERCISE 4.2.16. Let R be a commutative ring and F a free R-module with basis $\{b_1,\ldots,b_n\}$. Prove that if J is a proper ideal of R and $\pi : F \to F/JF$ is the natural homomorphism, then F/JF is a free R/J-module with basis $\{\pi(b_1),\ldots,\pi(b_n)\}$.

EXERCISE 4.2.17. Let R be a commutative ring and F a finitely generated free R-module. Show that any two bases for F have the same number of elements. (Hint: By Proposition 3.2.26, R contains a maximal ideal. Let $\mathfrak{m}$ be a maximal ideal in R and consider $F/\mathfrak{m}F$ as a vector space over $R/\mathfrak{m}$.)

EXERCISE 4.2.18. Let R be a commutative ring and $f \in R[x]$ a monic polynomial of degree n. Show that $S = R[x]/(f)$ is a free R-module of rank n and the set $\{1,x,x^2,\ldots,x^{n-1}\}$ is a free basis.

EXERCISE 4.2.19. Let R_1 and R_2 be rings and $R = R_1 \oplus R_2$.

(1) If M_1 and M_2 are left R_1 and R_2-modules respectively, show how to make $M_1 \oplus M_2$ into a left R-module.
(2) If M is a left R-module, show that there are R-submodules M_1 and M_2 of M such that $M = M_1 \oplus M_2$ and for each i, M_i is a left R_i-module.

EXERCISE 4.2.20. Let G be a group and H a subgroup. For any commutative ring R, let $\theta : R(H) \to R(G)$ be the homomorphism of rings induced by the set inclusion man $H \to G$ (see Example 3.2.4 (3)). Show that $R(G)$ is a free $R(H)$-module.

EXERCISE 4.2.21. Let V be a finitely generated vector space over a division ring D. Let $X \subseteq V$ be a spanning set for V. Show that there is a subset of X that is a basis for V. Do not assume X is finite.

EXERCISE 4.2.22. Let V be a vector space over a division ring D. Suppose there exists a positive number n such that every linearly independent subset of V has cardinality less than or equal to n. Show that V is finitely generated and $\dim_D(V) \leq n$.

EXERCISE 4.2.23. Let D be a division ring and V a nonzero vector space over D. As in Definition 4.1.27, a subset $X \subseteq V$ is said to be *linearly independent*, if every finite subset of X is linearly independent. We say X is a *basis* for V if X is a linearly independent spanning set for V. Apply Zorn's Lemma (Proposition 1.3.3) to prove the following.

(1) Every linearly independent subset of V is contained in a basis for V.
(2) If $S \subseteq V$ is a spanning set for V, then S contains a basis for V.

EXERCISE 4.2.24. Let D be a division ring and V a vector space over D. Let A and B be finite dimensional subspaces of V. Prove:

(1) $A+B$ is finite dimensional.
(2) $\dim_D(A+B) = \dim_D(A) + \dim_D(B) - \dim_D(A \cap B)$. (Hint: Apply Exercise 4.2.9 and Corollary 4.1.13 (1).)

3. Finitely Generated Modules over a Euclidean Domain

In Theorem 2.8.6 we proved that a finite abelian group is equal to the internal direct sum of cyclic subgroups. Every abelian group is a $\mathbb{Z}$-module and cyclic subgroups correspond to cyclic submodules. Therefore, we have already proved that every finite $\mathbb{Z}$-module is equal to the internal direct sum of cyclic submodules.

The proof given below of Theorem 4.3.1 uses a method that is commonly known as an "Artin Trick". For example, it is the proof used by [**7**, Theorem 4.5.1].

THEOREM 4.3.1. *Let M be a finitely generated $\mathbb{Z}$-module. Then there exists a finite subset $\{x_1,\dots,x_n\}$ of M such that $M = \mathbb{Z}x_1 \oplus \cdots \oplus \mathbb{Z}x_n$ is the internal direct sum of the cyclic submodules $\mathbb{Z}x_i$.*

PROOF. Assume M is a nonzero finitely generated $\mathbb{Z}$-module. If $\text{Rank}(M) = 1$, then M is cyclic and there is nothing to prove. The proof is by induction on q. Inductively assume $\text{Rank}(M) = q > 1$ and the theorem is true for all $\mathbb{Z}$-modules of rank less than q. The rest of the proof consists of a series of seven steps.

Step 0: This step reduces to the case where M is not free. Let $X = \{x_1,\dots,x_q\}$ be a generating set of M. As in Definition 4.1.24, by $\mathbb{Z}^{(q)}$ we denote the direct sum of q copies of the infinite cyclic group $\mathbb{Z}$. The function $\Sigma_X : \mathbb{Z}^{(q)} \to M$ defined by $\Sigma_X(r_1,\dots,r_q) = r_1x_1 + \cdots + r_qx_q$ is a $\mathbb{Z}$-module homomorphism. In fact, Σ_X is onto because X is a generating set for M. If there exists a finite generating set X for M such that the map Σ_X is an isomorphism, then we are done, because $\mathbb{Z}^{(q)}$ is a direct sum of cyclic submodules. Therefore, we assume that for every minimal generating set $X = \{x_1,\dots,x_q\}$, the kernel of Σ_X is nontrivial. That is, there exists $(r_1,\dots,r_q) \in \mathbb{Z}^{(q)}$ such that

$$0 = r_1x_1 + \cdots + r_qx_q \tag{3.1}$$

and $r_i \neq 0$ for some i. Notice that in this case there exists a relation (3.1) such that $r_i > 0$.

Step 1: Out of all minimal generating sets $x_1,\dots,x_q$ and all relations of the form (3.1), there is a least positive integer occurring as a coefficient r_i of some x_i. Pick one such generating set, say $a_1,\dots,a_q$, assume

$$0 = s_1a_1 + \cdots + s_qa_q, \tag{3.2}$$

and $s_1 > 0$ is minimal among all such positive coefficients in all such relations.

Step 2: We prove that if

$$0 = r_1a_1 + \cdots + r_qa_q, \tag{3.3}$$

then $s_1 \mid r_1$. By the division algorithm, $r_1 = s_1u + v$, where $0 \leq v < s_1$. Multiply (3.2) by u and subtract from (3.3) to get

$$0 = va_1 + (r_2 - s_2u)a_2 \cdots + (r_q - s_qu)a_q. \tag{3.4}$$

By minimality of s_1 we conclude $v = 0$.

Step 3: We prove that $s_1 \mid s_2$. Dividing, $s_2 = s_1u + v$, where $0 \leq v < s_1$. Set $a' = a_1 + ua_2$ and consider $a', a_2, \dots, a_q = a_1 + ua_2, a_2, \dots, a_q$. This set also generates M and

$$\begin{aligned} s_1a' + va_2 + s_3a_3 + \cdots + s_1a_q &= s_1a_1 + (s_1ua_2 + va_2) + s_3a_3 + \cdots + s_1a_q \\ &= s_1a_1 + \cdots + s_qa_q \\ &= 0. \end{aligned}$$

By minimality of s_1 we conclude that $v = 0$. The same argument shows that $s_1 \mid s_i$ for $i = 1, 2, \dots, q$.

Step 4: Set $a'' = a_1 + (s_2/s_1)a_2 + \cdots + (s_q/s_1)a_q$ and consider $a'', a_2, \ldots, a_q$. This set also generates M and

$$s_1 a'' = s_1 a_1 + s_2 a_2 + \cdots + s_q a_q = 0.$$

Set $A = (a'')$ and $B = (a_2, \ldots, a_q)$.

Step 5: We prove that $M = A \oplus B$. We already have $M = A + B = (a'', a_2, \ldots, a_q)$. By Proposition 4.1.20, it is enough to show $A \cap B = (0)$. Suppose $r_1 a'' = r_2 a_2 + \cdots + r_q a_q$. Then $r_1 a_1 + r_1(s_2/s_1)a_2 + \cdots + r_1(s_q/s_1)a_q = r_2 a_2 + \cdots + r_q a_q$. By Step 2, $s_1 \mid r_1$, and by Step 4, $r_1 a'' = 0$. This proves $A \cap B = (0)$, hence $M = A \oplus B$.

Step 6: Since $\mathrm{Rank}(B) \leq q-1$, by the induction hypothesis B is an internal direct sum of cyclic submodules. Since A is a cyclic submodule, this proves the theorem. □

THEOREM 4.3.2. *Let R be a euclidean domain and M a finitely generated R-module. Then there exists a finite subset $\{x_1, \ldots, x_n\}$ of M such that $M = Rx_1 \oplus \cdots \oplus Rx_n$ is the internal direct sum of the cyclic submodules Rx_i.*

PROOF. Let $\delta : R - (0) \to \mathbb{N}$ be the norm function on R. After only a few modifications the proof of Theorem 4.3.1 applies to R. In Step 1, pick the nonzero coefficient $s_1 \in R$ such that $\delta(s_1)$ is minimal. In Steps 2 and 3, when applying the division algorithm, there exist $u, v \in R$ such that either $v = 0$, or $\delta(v) < \delta(s_1)$. □

DEFINITION 4.3.3. Let R be an integral domain and M an R-module. If $x \in M$, then we say x is a *torsion element* of M in case there exists a nonzero $r \in R$ such that $rx = 0$. If every element of M is torsion, then we say M is torsion. Since R is an integral domain, by Exercise 4.3.9 the set of all torsion elements in M is a submodule of M, which is denoted M_t. If $M_t = 0$, then we say M is *torsion free*.

DEFINITION 4.3.4. Let R be a PID, M an R-module and $x \in M$. The cyclic submodule generated by x is Rx. Define $\theta_x : R \to M$ by $\theta(r) = rx$. Then θ_x is an R-module homomorphism. Denote by I_x the kernel of θ_x. That is,

$$I_x = \{r \in R \mid rx = 0\}$$

which is an ideal in R, hence is principal. So $I_x = Ra$ and up to associates in R, a is uniquely determined by x. We call a the *order of* x. The image of θ_x is Rx and by Theorem 4.1.12, $Rx \cong R/(I_x) \cong R/Ra$.

DEFINITION 4.3.5. Let R be a UFD and M a finitely generated R-module. By Example 4.1.16, the left regular representation $\lambda : R \to \mathrm{Hom}_R(M, M)$ is a homomorphism of rings that maps $r \in R$ to $\ell_r : M \to M$, where $\ell_r(x) = rx$ is "left multiplication by r". Let π be a prime element in R and n a positive integer. The kernel of ℓ_{π^n} is contained in the kernel of $\ell_{\pi^{n+1}}$. Therefore the union

$$\begin{aligned} M(\pi) &= \bigcup_{n>0} \ker(\ell_{\pi^n}) \\ &= \{x \in M \mid \text{there exists } n > 0 \text{ such that } \pi^n x = 0\} \end{aligned}$$

is a submodule of M.

EXAMPLE 4.3.6. Suppose M is a finitely generated abelian group of rank n. Consider the cases that can arise when $n \leq 3$.

(1) If $n = 1$, then M is cyclic. There are two cases: $M \cong \mathbb{Z}$, or $M \cong \mathbb{Z}/r_1$, for some $r_1 > 1$.
(2) If $n = 2$, then there are three cases:

(a) $M \cong \mathbb{Z} \oplus \mathbb{Z}$, or
(b) $M \cong \mathbb{Z}/r_1 \oplus \mathbb{Z}$, where $1 < r_1$, or
(c) $M \cong \mathbb{Z}/r_1 \oplus \mathbb{Z}/r_2$, where $1 < r_1 \leq r_2$, and $r_1 \mid r_2$.

(3) If $n = 3$, then there are four cases:
(a) $M \cong \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$, or
(b) $M \cong \mathbb{Z}/r_1 \oplus \mathbb{Z} \oplus \mathbb{Z}$, where $1 < r_1$, or
(c) $M \cong \mathbb{Z}/r_1 \oplus \mathbb{Z}/r_2 \oplus \mathbb{Z}$, where $1 < r_1 \leq r_2$, and $r_1 \mid r_2$, or
(d) $M \cong \mathbb{Z}/r_1 \oplus \mathbb{Z}/r_2 \oplus \mathbb{Z}/r_3$, where $1 < r_1 \leq r_2 \leq r_3$, and $r_1 \mid r_2 \mid r_3$.

THEOREM 4.3.7. *(Basis Theorem – Invariant Factor Form) Let R be a euclidean domain and M a finitely generated R-module. The following are true.*

(1) $M = F \oplus M_t$, where F is a free submodule of finite rank. The rank of F is uniquely determined by M.
(2) There exist $x_1, \ldots, x_\ell \in M_t$, and $r_1, \ldots, r_\ell \in R$ satisfying:
(a) $M_t = Rx_1 \oplus \cdots \oplus Rx_\ell$ is the internal direct sum of the cyclic submodules Rx_i,
(b) $r_1 \mid r_2 \mid r_3 \mid \cdots \mid r_\ell$, and $Rx_i \cong R/(r_iR)$,
(c) the integer ℓ is uniquely determined by M, and up to associates in R, the elements r_i are uniquely determined by M.

The elements $r_1, \ldots, r_\ell$ are called the invariant factors *of M.*

PROOF OF THE EXISTENCE CLAIM IN THEOREM 4.3.7. The proof is a continuation of the proof of Theorem 4.3.2. If M is free, then by Exercise 4.2.17 M is uniquely determined by $\mathrm{Rank}(M)$. Assume M is not free. Then by Step 5, there exists $x_1 \in M_t$ such that:

(1) The cyclic submodule Rx_1 is a direct summand of M. That is, $M = Rx_1 \oplus B$ for some submodule B.
(2) There exists $r_1 \in R$ such that $Rx_1 \cong R/Rr_1$.
(3) If B is free, then we are done.
(4) If B is not free, and $x \in B$ is torsion, then r_1 divides the order of x.

By Mathematical Induction on $\mathrm{Rank}(M)$, there exist $x_2, \ldots, x_\ell \in B$, $r_2, \ldots, r_\ell \in R$, $r_2 \mid r_3 \mid \cdots \mid r_\ell$, a free R-submodule F such that $B = Rx_2 \oplus \cdots \oplus Rx_\ell \oplus F$, and $Rx_i \cong R/Rr_i$ for each i. □

THEOREM 4.3.8. *(Basis Theorem – Elementary Divisor Form) Let R be a euclidean domain and M a finitely generated R-module. In the notation established above, the following are true.*

(1) $M = F \oplus M_t$, where F is a free submodule of finite rank. The rank of F is uniquely determined by M.
(2) $M_t = \bigoplus_\pi M(\pi)$ where π runs through a finite set of primes in R.
(3) For each prime π such that $M(\pi) \neq 0$, there exists a basis $\{a_1, \ldots, a_m\}$ such that $M(\pi) = Ra_1 \oplus Ra_2 \oplus \cdots \oplus Ra_m$ where the order of a_i is equal to π^{e_i} and $e_1 \geq e_2 \geq \cdots \geq e_m$.
(4) M_t is uniquely determined by the primes π that occur in (2) and the integers e_i that occur in (3).

The prime powers π^{e_i} that occur are called the elementary divisors *of M.*

PROOF. (Existence.) By Theorem 4.3.7, $M_t = Rx_1 \oplus \cdots \oplus Rx_\ell$. and $Rx_i \cong R/(r_iR)$. By Exercise 4.3.12, $R/(r_iR)$ is a direct sum of cyclic modules of the form R/π^e where π runs through the primes that divide r_i.

(Uniqueness.) The details are left to the reader. For example, the proof of the Basis Theorem for Finite Abelian Groups given in Theorem 2.8.7 is a good starting point. □

PROOF OF THE UNIQUENESS CLAIM IN THEOREM 4.3.7. Use the uniqueness of the elementary divisors in Theorem 4.3.8 to prove the uniqueness of the invariant factors in Theorem 4.3.7. □

3.1. Exercises.

EXERCISE 4.3.9. Let R be an integral domain and M an R-module. Let M_t be the set of all torsion elements in M (see Definition 4.3.3). Show that M_t is a submodule of M.

EXERCISE 4.3.10. Let R be a PID. Show that every nonzero ideal of R is a free R-module of rank 1.

EXERCISE 4.3.11. Let R be a PID. Let π be an irreducible element of R, $e > 0$ and $A = R/(\pi^e)$. Prove:

(1) Every ideal in A is principal.
(2) A is a field if and only if $e = 1$.
(3) A is a local ring, the unique maximal ideal is generated by π.
(4) A has exactly $e+1$ ideals, namely: $(0) \subseteq (\pi^{e-1}) \subseteq \cdots \subseteq (\pi^2) \subseteq (\pi) \subseteq A$.

EXERCISE 4.3.12. Let R be a PID. Let $\pi_1, \ldots, \pi_n$ be irreducible elements of R that are pairwise nonassociates. Let $e_1, \ldots, e_n$ be positive integers. If $x = \pi_1^{e_1}\pi_2^{e_2}\cdots\pi_n^{e_n}$, and $A = R/(x)$, prove:

(1) The ideals in A correspond to the divisors of x. Including the two trivial ideals (0) and A, there are exactly $(e_1+1)(e_2+1)\cdots(e_n+1)$ ideals in A.
(2) A has exactly n maximal ideals, namely $(\pi_1), \ldots, (\pi_n)$.
(3) A is isomorphic to the direct sum of the local rings $\bigoplus_i R/(\pi_i^{e_i})$.

EXERCISE 4.3.13. (The abelian group $\mathbb{Q}/\mathbb{Z}$) This exercise is a continuation of Exercises 2.2.29 and 2.3.21. For any integer $r \geq 1$, let $\ell_r : \mathbb{Q}/\mathbb{Z} \to \mathbb{Q}/\mathbb{Z}$ be the left multiplication by r map. Prove the following.

(1) Show that ℓ_r is onto for all $r \geq 1$. We say $\mathbb{Q}/\mathbb{Z}$ is a divisible abelian group.
(2) $\mathbb{Q}/\mathbb{Z}$ is a torsion $\mathbb{Z}$-module.
(3) The kernel of ℓ_r is a cyclic group of order r.
(4) If H is a finite subgroup of $\mathbb{Q}/\mathbb{Z}$, then H is cyclic. (Hint: Exercise 2.8.10.)
(5) If H is a finite subgroup of $\mathbb{Q}/\mathbb{Z}$, then $(\mathbb{Q}/\mathbb{Z})/H$ is isomorphic to $\mathbb{Q}/\mathbb{Z}$.

EXERCISE 4.3.14. (The p-torsion subgroup of $\mathbb{Q}/\mathbb{Z}$) Let p be a prime number. As in Section 4.3, let

$$\mathbb{Q}/\mathbb{Z}(p) = \bigcup_{n>0} \ker(\ell_{p^n})$$

be the subgroup of $\mathbb{Q}/\mathbb{Z}$ consisting of all elements annihilated by some power of p. Some authors denote the group $\mathbb{Q}/\mathbb{Z}(p)$ by $\mathbb{Z}(p^\infty)$. Prove the following.

(1) Every proper subgroup of $\mathbb{Q}/\mathbb{Z}(p)$ is a finite cyclic group.
(2) $\mathbb{Q}/\mathbb{Z}(p)$ is a divisible group (see Exercise 4.3.13 (1)).
(3) $\mathbb{Q}/\mathbb{Z}$ is equal to the internal direct sum $\bigoplus_{p\in P} \mathbb{Q}/\mathbb{Z}(p)$, where P is the set of all prime numbers.
(4) If H is a proper subgroup of $\mathbb{Q}/\mathbb{Z}(p)$, then the quotient $\mathbb{Q}/\mathbb{Z}(p)/H$ is isomorphic to $\mathbb{Q}/\mathbb{Z}(p)$.

4. Algebras

DEFINITION 4.4.1. Let R be a commutative ring, A a ring and $\theta : R \to A$ a homomorphism of rings such that $\theta(R)$ is a subring of the center of A. Then we say A is an *R-algebra* and θ is the *structure homomorphism*. If A and B are two R-algebras, then an *R-algebra homomorphism from A to B* is a function $\phi : A \to B$ satisfying:

(1) $\phi(1) = 1$,
(2) $\phi(x+y) = \phi(x) + \phi(y)$ for all $x, y \in A$,
(3) $\phi(xy) = \theta(x)\phi(y)$ for all $x, y \in A$, and
(4) $\phi(a) = a$ for all $a \in k$.

An *R-algebra isomorphism from A to B* is a homomorphism $\phi : A \to B$ that is one-to-one and onto. An *R-algebra automorphism of A* is a homomorphism from A to A that is one-to-one and onto. The set of all R-algebra automorphisms is a group and is denoted $\mathrm{Aut}_R(A)$.

If k is a field and A is a k-algebra, then the structure homomorphism is necessarily one-to-one, so it is convenient to identify k as a subring of the center of A. In this case, A is a left k-vector space by virtue of the multiplication and addition operations on A.

EXAMPLE 4.4.2. Important examples of algebras over a field are listed here.

(1) If F and k are fields and k is a subfield of F, then we say F/k is an *extension of fields*. In this case F is a k-algebra.
(2) The ring of polynomials $k[x]$ is a k-algebra where we identify k with the constant polynomials. Because $1, x, x^2, \ldots$ are linearly independent over k, $\dim_k(k[x]) = \infty$.
(3) Let $q \in k[x]$ be a polynomial of degree $n > 0$. In Lemma 4.4.3 below we prove that the quotient ring $k[x]/(q)$ is a commutative k-algebra of dimension n.

Let k be a field, x an indeterminate, and q a polynomial in $k[x]$. The principal ideal generated by q is $(q) = \{fq \mid f \in k[x]\}$, which is equal to the set of all polynomials that are divisible by q. By 3.2.14, $k[x]/(q)$ is a commutative ring.

LEMMA 4.4.3. *In the above context, the following are true.*

(1) $k[x]/(q)$ is a commutative k-algebra.
(2) $k[x]/(q)$ is a k-vector space.
(3) $\dim_k(k[x]/(q)) = \begin{cases} \infty & \text{if } q = 0 \\ \deg q & \text{if } q \neq 0. \end{cases}$
(4) If $(q) \neq k[x]$, then $k[x]/(q)$ is a k-algebra.
(5) $k[x]/(q)$ is a field if and only if q is irreducible.

PROOF. Since k is a subring of $k[x]$, $k[x]$ is a k-algebra. If $q = 0$, then $k[x]/(q) = k[x]$ is not finite dimensional (Example 4.4.2 (3)). If $q \neq 0$ and $n = \deg q$, then by Exercise 4.2.18, $k[x]/(q)$ is a k-vector space and $\{[1], [x], \ldots, [x^{n-1}]\}$ is a k-basis for $k[x]/(q)$. Since $k[x]$ is a PID, $k[x]/(q)$ is a field if and only if q is irreducible, by Exercise 4.3.12. If $\deg q = 0$, then $k[x]/(q)$ is the trivial ring and is not a k-algebra. Otherwise, $k[x]/(q)$ is a k-algebra. □

DEFINITION 4.4.4. Let A be a k-algebra. If $X \subseteq A$, then by $k[X]$ we denote the k-subalgebra of A generated by k and X. Thus $k[X]$ is the smallest subring of A that contains both k and X.

DEFINITION 4.4.5. Let k be a field, A a k-algebra, and α an element of A. If there is a nonzero polynomial $f \in k[x]$ and $f(\alpha) = 0$, then we say α is *algebraic over k*. Otherwise

we say α is *transcendental over k*. We say A is *algebraic over k* if every $\alpha \in A$ is algebraic over k.

THEOREM 4.4.6. *(Fundamental Theorem on Algebraic Elements) Let k be a field, A a k-algebra, and $\alpha \in A - \{0\}$. There is a k-algebra homomorphism $\tau : k[x] \to A$ satisfying the following.*

(1) $\tau(x) = \alpha$.
(2) The kernel of τ is $I(\alpha) = \{p \in k[x] \mid p(\alpha) = 0\}$. There is a polynomial $f \in k[x]$ such that $I(\alpha)$ is equal to the principal ideal (f) generated by f.
(3) The image of τ is $k[\alpha]$, the subalgebra of A generated by k and α.
(4) α is transcendental over k if and only if $I(\alpha) = (0)$.
(5) α is algebraic over k if and only if $I(\alpha) \neq (0)$. In this case, $\deg f > 0$, $\dim_k k[\alpha] = \deg f$, *$f$ can be taken to be monic, and if $p \in I(\alpha)$, then $f \mid p$.*
(6) $k[\alpha] \cong k[x]/(f)$.
(7) $k[\alpha]$ is a commutative principal ideal ring.

The polynomial f is called the minimal polynomial of α *and is denoted* $\min.\mathrm{poly}_k(\alpha)$. *If α is algebraic and f is taken to be monic, then f is uniquely determined by α.*

PROOF. Given $\alpha \in A$, the evaluation homomorphism $\tau : k[x] \to A$, is a k-algebra homomorphism determined by $x \mapsto \alpha$ (Theorem 3.6.2). Since $k[x]$ is a principal ideal domain (Corollary 3.6.5), there exists a polynomial $f \in k[x]$ which generates the kernel of τ. The image of τ is denoted $k[\alpha]$. By Exercise 3.6.33, $k[\alpha]$ is a commutative principal ideal ring and is the smallest subring of A containing k and α. By Proposition 3.2.15, $k[\alpha] \cong k[x]/(f)$. By Definition 4.4.5, α is transcendental if and only if $I(\alpha) = (0)$. In this case, τ is one-to-one and $k[\alpha] \cong k[x]$. If $I(\alpha) \neq (0)$, then $\deg f \geq 1$ and f is unique up to associates in $k[x]$. Hence if f is taken to be monic, then f is unique. Let $f = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ be the minimal polynomial of α, where $n \geq 1$. Exercise 4.2.18 says $k[\alpha]$ is a k-vector space of dimension n spanned by $1, \alpha, \ldots, \alpha^{n-1}$. □

EXAMPLE 4.4.7. If x is an indeterminate, and $k(x)$ is the field of rational functions over k, then $k[x] \to k(x)$ is one-to-one (Lemma 3.5.1). Hence x is transcendental over k.

COROLLARY 4.4.8. *If k is a field and A is a finite dimensional k-algebra, then A is algebraic over k. If $\alpha \in A$ and* $\dim_k(A) = n$, *then the degree of* $\min.\mathrm{poly}_k(\alpha)$ *is less than or equal to n.*

PROOF. Let $\alpha \in A$, and $\dim_k(A) = n$. By Theorem 4.2.4, the set $\{u^n, u^{n-1}, \ldots, u, 1\}$ is linearly dependent. A dependence relation $0 = a_n u^n + a_{n-1}u^{n-1} + \cdots + a_1 u + a_0$ over k shows that u is algebraic over k. □

COROLLARY 4.4.9. *Let k be a field and A a k-algebra. If $\alpha \in A$ is algebraic over k, then $k[\alpha]$ is algebraic over k.*

PROOF. By Theorem 4.4.6 (5), $k[\alpha]$ is finite dimensional over k. □

COROLLARY 4.4.10. *Let k be a field, A a k-algebra, and u an element of A that is algebraic over k. Then u is an invertible element of A if and only if* $\min.\mathrm{poly}_k(u)$ *has a nonzero constant term.*

PROOF. Let $f(x) = \min.\mathrm{poly}_k(u) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. If $u \in k$, then $f(x) = x - u$ and in this case the result holds. Assume $n \geq 2$. We have $f(u) = u^n + a_{n-1}u^{n-1} + \cdots + a_1 u + a_0 = 0$. Solving for a_0 and factoring, we get

$$-a_0 = u(u^{n-1} + a_{n-1}u^{n-2} + \cdots + a_1). \tag{4.1}$$

Assume $a_0 = 0$ and for sake of contradiction assume u is invertible. Then multiplying (4.1) by u^{-1} on both sides we get $u^{n-1} + a_{n-1}u^{n-2} + \cdots + a_1 = 0$, which contradicts the definition of the minimal polynomial of u in Theorem 4.4.6. Conversely, assume $a_0 \neq 0$. Multiplying both sides of (4.1) by $-a_0^{-1}$, we get

$$1 = u(-a_0^{-1})(u^{n-1} + a_{n-1}u^{n-2} + \cdots + a_1)$$

which shows u is invertible in A. □

THEOREM 4.4.11. *Let k be a field, A a finite dimensional k-algebra, and $u \in A$.*

(1) If u is not a zero divisor, then u is invertible.
(2) If A is a domain (that is, A has no zero divisors), then A is a division ring.

PROOF. By Corollary 4.4.8, A is algebraic over k. The proof is by contraposition. Assume A contains a nonzero element u which is not invertible. We show that u is a zero divisor in A. Let $f = \min.\text{poly}_k(u) \in k[x]$. By Corollary 4.4.10, u is invertible if and only if $f(0) \in k - (0)$. Assume $f(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x$ has zero constant term. By Eq. (4.1),

$$0 = u(u^{n-1} + a_{n-1}x^{n-2} + \cdots + a_1).$$

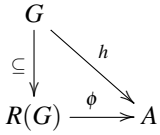
Since the minimum polynomial for u has degree n, we know $u^{n-1} + a_{n-1}x^{n-2} + \cdots + a_1 \neq 0$. This shows u is a zero divisor in A. □

4.1. Exercises.

EXERCISE 4.4.12. Let R be a commutative ring and A an R-algebra. Suppose $\alpha \in A$ is a root of the polynomial $p \in R[x]$. Prove:

(1) If B is another R-algebra and $\phi : A \to B$ is an R-algebra homomorphism, then $\phi(\alpha)$ is a root of p.
(2) If u is a unit in A, then $u^{-1}\alpha u$ is a root of p.

EXERCISE 4.4.13. (Universal Mapping Property) Let R be a commutative ring, G a finite group, and $R(G)$ the group ring (see Example 3.1.6). Let A be an R-algebra and $h : G \to A^*$ a homomorphism from G to the group of units of A. Show that there is a unique homomorphism of R-algebras $\phi : R(G) \to A$ such that diagram



commutes. Show that the same result holds if G is a group that is not necessarily finite.

EXERCISE 4.4.14. Let R be a commutative ring and M a finitely generated free R-module of rank n. Using Exercises 4.1.41 and 4.4.13, show that there exists an R-algebra homomorphism $\phi : R(S_n) \to \text{Hom}_R(M,M)$ from the group ring to the ring of endomorphisms. Show that in general ϕ is not one-to-one.

5. Matrix Theory

If R is a commutative ring and M and N are finitely generated free modules over R, then we show that any R-module homomorphism $\phi : M \to N$ can be represented as a matrix. The matrix representation of ϕ depends on a choice of bases for M and N. When M is free of rank n, then we show that there is an isomorphism of R-algebras $\text{Hom}_R(M,M) \cong M_n(R)$.

5.1. The Matrix of a Linear Transformation.

DEFINITION 4.5.1. Let R be any ring and m,n positive integers. By $M_{nm}(R)$ we denote the set of all n-by-m matrices over R. If $m = n$, then we simply write $M_n(R)$ instead of $M_{nn}(R)$. Addition of matrices is coordinate-wise $(\alpha_{ij}) + (\beta_{ij}) = (\alpha_{ij} + \beta_{ij})$. We can multiply by elements of R from the left $r(\alpha_{ij}) = (r\alpha_{ij})$. If $(\alpha_{ij}) \in M_{nm}(R)$ and $(\beta_{jk}) \in M_{mp}(R)$, then the matrix product is defined by $(\alpha_{ij})(\beta_{jk}) = (\gamma_{ik}) \in M_{np}(R)$, where $\gamma_{ik} = \sum_{j=1}^{m} \alpha_{ij}\beta_{jk}$. If R is a commutative ring, in Corollary 4.5.7 below we prove that $M_n(R)$ is an R-algebra. When R is an arbitrary ring, see [**4**, Section 4.3.1] for the proof that $M_n(R)$ is a ring that contains R as a subring.

DEFINITION 4.5.2. Let e_{ij} be the matrix with 1 in position (i, j) and 0 elsewhere. The matrix e_{ij} is called an *elementary* matrix (see Example 3.2.10).

LEMMA 4.5.3. *For a ring R, the set $M_{nm}(R)$ of n-by-m matrices over R is a free R-module. The set $\{e_{ij} \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ of elementary matrices is a free basis with nm elements.*

PROOF. See Definition 4.1.24 for the definition of free module. The proof is left to the reader. □

DEFINITION 4.5.4. Let R be any ring, M a free R-module of rank m and N a free R-module of rank n. Let $X = \{x_1, \ldots, x_m\}$ be a basis for M and $Y = \{y_1, \ldots, y_n\}$ a basis for N. Given $\phi \in \mathrm{Hom}_R(M,N)$, ϕ maps $x_j \in X$ to a linear combination of Y. That is,

$$\phi(x_j) = \sum_{i=1}^{n} \phi_{ij} y_i$$

where the elements ϕ_{ij} are in R. The *matrix of ϕ with respect to the bases X and Y* is defined to be $M(\phi, X, Y) = (\phi_{ij})$, which is a matrix in $M_{nm}(R)$.

PROPOSITION 4.5.5. *Let R be any ring. If M is a free R-module of rank m, and N is a free R-module of rank n, then there is a $\mathbb{Z}$-module isomorphism $\mathrm{Hom}_R(M,N) \cong M_{nm}(R)$. If R is a commutative ring, then this is an R-module isomorphism and $\mathrm{Hom}_R(M,N)$ is a free R-module of rank mn.*

PROOF. Let $X = \{x_1, \ldots, x_m\}$ be a basis for M and $Y = \{y_1, \ldots, y_n\}$ a basis for N. The assignment $\phi \mapsto M(\phi, X, Y)$ defines a $\mathbb{Z}$-module homomorphism

$$M(\cdot, X, Y) : \mathrm{Hom}_R(M,N) \to M_{nm}(R).$$

Conversely, if $(\alpha_{ij}) \in M_{nm}(R)$, define α in $\mathrm{Hom}_R(M,N)$ by

$$\alpha(x_j) = \sum_{i=1}^{n} \alpha_{ij} y_i.$$

The rest is left to the reader. □

PROPOSITION 4.5.6. *Let R be any ring. Let M, N, and P denote free R-modules, each of finite rank. Let X, Y and Z be bases for M, N, and P respectively. Let $\phi \in \mathrm{Hom}_R(M,N)$ and $\psi \in \mathrm{Hom}_R(N,P)$. If the matrices $M(\psi, Y, Z)$ and $M(\phi, X, Y)$ are treated as having entries from the ring R^o, the opposite ring of R, then*

$$M(\psi\phi, X, Z) = M(\psi, Y, Z)M(\phi, X, Y).$$

PROOF. The opposite ring R^o is defined as in Definition 3.1.8. Let $X = \{x_1, \ldots, x_m\}$, $Y = \{y_1, \ldots, y_n\}$, and $Z = \{z_1, \ldots, z_p\}$. Let $M(\phi, X, Y) = (\phi_{ij})$, $M(\psi, Y, Z) = (\psi_{ij})$. It follows from

$$\psi\phi(x_j) = \psi\left(\sum_{i=1}^{n} \phi_{ij} y_i\right) = \sum_{i=1}^{n} \phi_{ij} \sum_{k=1}^{p} \psi_{ki} z_k = \sum_{k=1}^{p} \left(\sum_{i=1}^{n} \phi_{ij}\psi_{ki}\right) z_k$$

that $M(\psi\phi, X, Z) = (\gamma_{kj})$, where $\gamma_{kj} = \sum_{i=1}^{n} \phi_{ij}\psi_{ki}$. Computing the product of the two matrices over R^o, we get $M(\psi, Y, Z)M(\phi, X, Y) = (\tau_{kj})$, where

$$\tau_{kj} = \sum_{i=1}^{n} \psi_{ki} * \phi_{ij} = \sum_{i=1}^{n} \phi_{ij}\psi_{ki}.$$

□

COROLLARY 4.5.7. *Let R be any ring. With the binary operations defined in Definition 4.5.1, $M_n(R)$ is a ring with identity element $I_n = e_{11} + \cdots + e_{nn}$. The set $R \cdot I_n$ of all scalar matrices in $M_n(R)$ is a subring which is isomorphic to R. The center of the ring $M_n(R)$ is equal to the center of the subring $R \cdot I_n$. If R is commutative, the matrix ring $M_n(R)$ is an R-algebra and the center of $M_n(R)$ is equal to $R \cdot I_n$.*

PROOF. Use Proposition 4.5.6 to show that matrix multiplication is associative. If R is commutative, as shown in Example 3.1.13, the center of $M_n(R)$ is equal to the set of scalar matrices. The same proof can be used to prove that the center of $M_n(R)$ is equal to the center of the subring $R \cdot I_n$. The rest is left to the reader. □

PROPOSITION 4.5.8. *Let R be any ring. If M is a free R-module of rank n, then there is an isomorphism of rings $\mathrm{Hom}_R(M, M) \cong M_n(R^o)$. If R is commutative, this is an isomorphism of R-algebras.*

PROOF. Pick a basis for M. The map of Proposition 4.5.5 defines an isomorphism of abelian groups. It is multiplicative by Proposition 4.5.6. □

DEFINITION 4.5.9. Let R be a commutative ring and $n \geq 1$. If A, B are matrices in $M_n(R)$ and P is an invertible matrix in $M_n(R)$ such that $A = P^{-1}BP$, then we say A and B are *similar*. The reader should verify that this defines an equivalence relation on $M_n(R)$.

PROPOSITION 4.5.10. *Let R be a commutative ring and M a free R-module of rank n. Let X and Y be two bases for M. If $\phi \in \mathrm{Hom}_R(M, M)$, then the matrix $M(\phi, X, X)$ of ϕ with respect to X and the matrix $M(\phi, Y, Y)$ of ϕ with respect to Y are similar. In fact, if $1 \in \mathrm{Hom}_R(M, M)$ is the identity map, then $M(1, X, Y)^{-1} = M(1, Y, X)$ and $M(\phi, X, X) = M(1, Y, X)M(\phi, Y, Y)M(1, X, Y)$.*

PROOF. Let $I \in M_n(R)$ be the identity matrix. It follows from Proposition 4.5.6 that $I = M(1, X, X) = M(1, Y, Y)$, $M(1, X, Y)M(1, Y, X) = I$, and $M(1, Y, X)M(1, X, Y) = I$. Also $M(\phi, X, Y) = M(1, X, Y)M(\phi, X, X) = M(\phi, Y, Y)M(1, X, Y)$. □

EXAMPLE 4.5.11. Let R be a commutative ring and $A \in M_{mn}(R)$. Elements of R^n can be viewed as n-by-1 column matrices in M_{n1}. As in Proposition 4.5.5, multiplication by A from the left defines an element in $\mathrm{Hom}_R(R^n, R^m)$. In particular, if k is a field and $A \in M_n(k)$, then left multiplication by A defines a linear transformation from k^n to k^n. We define the rank of A and the nullity of A as in Exercise 4.2.11. Define the *column space* of A to be the subspace of k^n spanned by the columns of A. The rank of A is seen to be the dimension of the column space of A.

5.2. The Transpose of a Matrix and the Dual of a Module.

DEFINITION 4.5.12. Let R be a commutative ring. Let M be a left R-module. The *dual of* M is defined to be $M^* = \mathrm{Hom}_R(M,R)$. We turn M^* into a right R-module by the action $(fr)(x) = (f(x))r$, for $r \in R$, $f \in M^*$, $x \in M$. The reader should verify that this is a well defined right R-module action on M^*. If N is another left R-module, and $\psi \in \mathrm{Hom}_R(M,N)$, define $\psi^* : N^* \to M^*$ by the rule $\psi^*(f) = f \circ \psi$, for any $f \in N^*$.

LEMMA 4.5.13. *Let R be a commutative ring. Let M and N be left R-modules. If $\psi : M \to N$ is a homomorphism of left R-modules, then $\psi^* : N^* \to M^*$ is a homomorphism of right R-modules. If L is another R-module, and $\phi \in \mathrm{Hom}_R(L,M)$, then $(\psi\phi)^* = \phi^*\psi^*$.*

PROOF. Let $f, g \in N^*$ and $a \in R$. The reader should verify that $\psi^*(f+g) = \psi^*(f) + \psi^*(g)$. If $x \in M$, then

$$(\psi^*(fa))(x) = (fa)(\psi(x)) = (f(\psi(x)))a = (\psi^*(f)(x))a = (\psi^*(f)a)(x).$$

Lastly, $\phi^*\psi^*(f) = (\psi\phi)^*(f)$. □

DEFINITION 4.5.14. Let R be a commutative ring. Let M be a left R-module which is free of finite rank. If $B = \{v_1, \ldots, v_n\}$ is a basis for M, then define $v_1^*, \ldots, v_n^*$ in M^* by the rules

$$v_i^*(v_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

PROPOSITION 4.5.15. *Let R be a commutative ring. If M is a free left R-module with basis $B = \{v_1, \ldots, v_n\}$, then M^* is a free right R-module with basis $B^* = \{v_1^*, \ldots, v_n^*\}$.*

PROOF. By Proposition 4.5.5, M^* is isomorphic to $M_{1n}(R)$ as $\mathbb{Z}$-modules. Under this isomorphism, v_i^* is mapped to the row matrix e_{1i} which has 1 in position i and zeros elsewhere. This is therefore a homomorphism of right R-modules. □

THEOREM 4.5.16. *Let R be a commutative ring. Let M and N be free R-modules, each of finite rank. Let X be a basis for M, and Y a basis for N. Let X^* and Y^* be the corresponding bases for M^* and N^*. Given $\phi \in \mathrm{Hom}_R(M,N)$,*

$$M(\phi^*, Y^*, X^*) = M(\phi, X, Y)^T.$$

That is, the matrix of ϕ^* with respect to Y^* and X^* is the transpose of the matrix of ϕ with respect to X and Y.

PROOF. Let $X = \{u_1, \ldots, u_m\}$ and $Y = \{v_1, \ldots, v_n\}$. Let $M(\phi, X, Y) = (\phi_{ij})$. Consider $\phi^*(v_l^*)(u_j) = v_l^*(\phi(u_j)) = v_l^*(\sum_{i=1}^n \phi_{ij} v_i) = \phi_{lj}$. Now consider $(\sum_{i=1}^m \phi_{li} u_i^*)(u_j) = \phi_{lj}$. Therefore, $\phi^*(v_l^*) = \sum_{i=1}^m \phi_{li} u_i^*$ as elements of $M^* = \mathrm{Hom}_R(M,R)$ because they agree on a basis of M. This also shows that column l of the matrix $M(\phi^*, Y^*, X^*)$ is the transpose of $(\phi_{l1}, \phi_{l2}, \ldots, \phi_{lm})$, which is row l of $M(\phi, X, Y)$ □

DEFINITION 4.5.17. If k is a field, the space $V^{**} = \mathrm{Hom}_k(V^*, k)$ is called the *double dual* of V. Given $v \in V$, let $\varphi_v : V^* \to k$ be the "evaluation at v" map. That is, if $f \in V^*$, then $\varphi_v(f) = f(v)$. The reader should verify that φ_v is an element of V^{**}, and that the assignment $v \mapsto \varphi_v$ is a homomorphism of k-vector spaces $V \to V^{**}$.

THEOREM 4.5.18. *Let V be a finitely generated vector space over a field k. The map $V \to V^{**}$ which sends a vector $v \in V$ to φ_v is a vector space isomorphism.*

PROOF. Let v be a nonzero vector in V. By Theorem 4.2.4, we can extend $\{v\}$ to a basis for V, say $B = \{v, v_2, \ldots, v_n\}$. Define $f \in V^*$ to be the projection mapping onto the v-coordinate. Then $f(v) = 1$, and $f(v_i) = 0$ for $2 \leq i \leq n$. Then $\varphi_v(f) = f(v) = 1$. This proves $V \to V^{**}$ is one-to-one. If V is finite dimensional, then $V \to V^{**}$ is onto since $\dim_k(V) = \dim_k(V^{**})$. □

Theorem 4.5.18 extends to finitely generated projective modules over any ring (see [**4**, Exercise 6.5.22]).

THEOREM 4.5.19. *Let D be a field and V and W finitely generated D-vector spaces. Let $\phi \in \mathrm{Hom}_D(V,W)$. Let $\phi^* : W^* \to V^*$ be the associated homomorphism of right D-vector spaces.*

(1) If ϕ is one-to-one, then ϕ^* is onto.
(2) If ϕ is onto, then ϕ^* is one-to-one.
(3) The rank of ϕ is equal to the rank of ϕ^*.

PROOF. (1): Assume ϕ is one-to-one. Let $f : V \to D$ be in V^*. By Exercise 4.2.15 there is $\bar{f} : W \to D$ in W^* such that $f = \bar{f}\phi = \phi^*(\bar{f})$.

(2): Assume ϕ is onto. A typical element of W is of the form $w = \phi(v)$, for some $v \in V$. Assume $g \in W^*$ and $g\phi = 0$. Then $g(w) = g(\phi(v)) = 0$.

(3): Let $n = \dim_D(V)$. By Proposition 4.5.5, $\dim_D(V^*) = n$. Let $U = \ker\phi$. Let $\psi : U \to V$ be the inclusion map. By (1), ψ^* is onto. Then $\mathrm{Rank}(\psi^*) = \dim(U^*) = \dim(U) = \mathrm{Nullity}(\phi) = n - \mathrm{Rank}\,\phi$. By Lemma 4.5.13, $\mathrm{im}\,\phi^* \subseteq \ker\psi^*$. We prove the reverse inclusion. Suppose $f \in V^*$ and $\psi^*(f) = f\psi = 0$. Then f factors through $V/\ker\phi = \mathrm{im}\,\phi$. There is $\bar{f} : \mathrm{im}\,\phi \to D$ such that $f = \bar{f}\phi$. By Exercise 4.2.15, $\bar{f}$ extends to W, so f is in the image of ϕ^*. This proves $\mathrm{Rank}\,\phi^* = \mathrm{Nullity}\,\psi^* = n - \mathrm{Rank}\,\psi^* = \mathrm{Rank}\,\phi$. □

COROLLARY 4.5.20. *Let D be a field and $A \in M_{nm}(D)$. The row rank of A is equal to the column rank of A.*

PROOF. As in Proposition 4.5.5, define α in $\mathrm{Hom}_D(D^m, D^n)$ to be "left multiplication by A". Let α^* be the associated map on dual spaces. By Theorem 4.5.16 the matrix of α^* is A^T. The column rank of A is equal to $\mathrm{Rank}\,\alpha$ which is equal to $\mathrm{Rank}\,\alpha^*$, by Theorem 4.5.19. But $\mathrm{Rank}\,\alpha^*$ is equal to the column rank of A^T, which is the row rank of A. □

5.3. Exercises.

EXERCISE 4.5.21. Let k be a field and V a finite dimensional vector space over k. Show that $\mathrm{Hom}_k(V,V)$ is a commutative ring if and only if $\dim_k(V) \leq 1$.

EXERCISE 4.5.22. Suppose $\phi \in \mathrm{Hom}_D(V,V)$, where V is a finite dimensional vector space over the field D. Prove:

(1) There is a chain of subspaces $\ker(\phi) \subseteq \ker(\phi^2) \subseteq \ker(\phi^3) \subseteq \cdots$.
(2) There is a chain of subspaces $\phi(V) \supseteq \phi^2(V) \supseteq \phi^3(V) \supseteq \cdots$.
(3) The kernel of $\phi : \phi(V) \to \phi^2(V)$ is equal to $\ker(\phi) \cap \phi(V)$. More generally, if $m \geq 1$, the kernel of $\phi^m : \phi^m(V) \to \phi^{2m}(V)$ is equal to $\ker(\phi^m) \cap \phi^m(V)$.
(4) If $m \geq 1$ and $\phi^m(V) = \phi^{m+1}(V)$, then $\phi^m(V) = \phi^{m+i}(V)$ for all $i \geq 1$.
(5) If $n = \dim_D(V)$, then there exists m such that $1 \leq m \leq n$ and $\phi^m(V) = \phi^{m+1}(V)$.
(6) If $n = \dim_D(V)$, then there exists m such that $1 \leq m \leq n$ and $\ker(\phi^m) \cap \phi^m(V) = (0)$.

EXERCISE 4.5.23. Let R be a commutative ring. Let $A \in M_{nm}(R)$ and $B, C \in M_{ml}(R)$. Prove:

(1) $(A^T)^T = A$.
(2) $(B+C)^T = B^T + C^T$.
(3) $(AB)^T = B^T A^T$.

EXERCISE 4.5.24. If R is a commutative ring, show that the mapping $M_n(R) \to M_n(R)^o$ defined by $A \mapsto A^T$ is an isomorphism of R-algebras.

EXERCISE 4.5.25. If R is any ring, show that the mapping $M_n(R) \to M_n(R^o)^o$ defined by $A \mapsto A^T$ is an isomorphism of rings. Using the Morita Theorems, a very general version of this is proved in [**4**, Corollary 6.9.3].

EXERCISE 4.5.26. Let R be any ring, M and N finitely generated R-modules, and $\phi \in \mathrm{Hom}_R(M,N)$. Show that there exist positive integers m and n, epimorphisms $f : R^{(m)} \to M$, $g : R^{(n)} \to N$, and $\theta \in \mathrm{Hom}_R(R^{(m)}, R^{(n)})$ such that the diagram

$$\begin{array}{ccc} R^{(m)} & \xrightarrow{\theta} & R^{(n)} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\ M & \xrightarrow{\phi} & N \end{array}$$

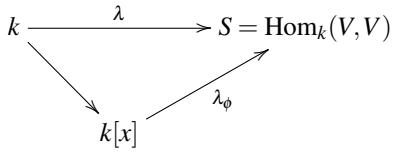
commutes. Therefore, given generators for M and N, ϕ can be represented as a matrix.

5.4. The Canonical Form of a Linear Transformation. If k is a field, V a finite dimensional k-vector space, and $\phi : V \to V$ a linear transformation, then we show that there is basis for V such that the matrix of ϕ is in so-called rational canonical form. The method of proof is to make V into a module over the polynomial ring $k[x]$ using ϕ and apply Theorem 4.3.7. Assuming the minimal polynomial for ϕ splits over k, we show that there is a basis for V such that the matrix of ϕ is in so-called Jordan canonical form. The proof is an application of Theorem 4.3.8. With respect to the standard basis, a matrix in $M_n(k)$ defines a linear transformation on $k^{(n)}$. By treating a matrix A as a linear transformation, we define the rational canonical form for A. The canonical form is a unique matrix in the similarity class containing A. Two matrices are similar if and only if they have the same canonical form.

5.4.1. *A vector space as a $k[\phi]$-module.* Let k be a field and V a k-vector space. Let $S = \mathrm{Hom}_k(V,V)$. By Proposition 4.5.5, S is finite dimensional as a k-vector space. By Corollary 4.4.8, S is algebraic over k. By Theorem 4.4.6, every ϕ in $S = \mathrm{Hom}_k(V,V)$ has a minimal polynomial $f = \mathrm{min.poly}_k(\phi)$. By Proposition 4.5.8, $M_n(k)$ and $\mathrm{Hom}_k(V,V)$ are isomorphic as k-algebras. If X is a basis for V, and $A = M(\phi, X, X)$, then f equal to both $\mathrm{min.poly}_k(\phi)$ and $\mathrm{min.poly}_k(A)$. The evaluation homomorphism $\lambda_\phi : k[x] \to \mathrm{Hom}_k(V,V)$ (Theorem 3.6.2) defined by $x \mapsto \phi$ maps $k[x]$ onto the commutative subring $k[\phi]$. There is a k-algebra isomorphism $k[x]/(f) \cong k[\phi]$ (Theorem 4.4.6). Since $k[x]$ is a PID, by Corollary 3.2.17 every ideal in $k[\phi]$ is a principal ideal. The ideals in $k[\phi]$ correspond up to associates to the divisors of f in $k[x]$ (see Exercises 4.3.11 and 4.3.12).

By Exercise 4.1.34, V is a left S-module by the action $\psi v = \psi(v)$, for any $\psi \in S$ and $v \in V$. By Example 4.1.16, the left regular representation $\lambda : k \to S = \mathrm{Hom}_k(V,V)$ is a homomorphism of rings that maps k into the center of S. Since k is a field, this map is one-to-one. Let $\phi \in \mathrm{Hom}_k(V,V)$. Using this ϕ, we make V into a left $k[x]$-module. By Theorem 3.6.2, the evaluation homomorphism $\lambda_\phi : k[x] \to S$ which maps x to ϕ is a

homomorphism of rings.



If $p(x) = a_0 + a_1x + \cdots + a_nx^n \in k[x]$, then $\lambda_\phi(p(x)) = a_0 + a_1\phi + \cdots + a_n\phi^n$. The image of λ_ϕ is the commutative k-algebra denoted $k[\phi]$. By Example 4.1.4 (4), λ_ϕ turns V into a $k[x]$-module. For any $v \in V$ and $p(x) \in k[x]$, the left multiplication of v by $p(x)$ is given by the formula:

$$\begin{aligned} p(x)v &= \lambda_\phi(p(x))v \\ &= (a_0 + a_1\phi + \cdots + a_n\phi^n)v \\ &= a_0v + a_1\phi(v) + \cdots + a_n\phi^n(v). \end{aligned}$$

By V_ϕ we denote the left $k[x]$-module structure on V induced by λ_ϕ. Since V is finitely generated as a k-vector space, V_ϕ is finitely generated as a $k[x]$-module. By Corollary 3.6.5, $k[x]$ is a euclidean domain. By Theorem 4.3.2, V_ϕ is the internal direct sum of cyclic submodules. That is, there exist $v_1, \ldots, v_q$ in V such that $V_\phi = (v_1) \oplus \cdots \oplus (v_q)$, where $(v_i) = k[\phi]v_i$.

If $f = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$ is the minimal polynomial of ϕ, then a k-basis for $k[\phi]$ is $\{\phi^{n-1}, \ldots, \phi, 1\}$ (Theorem 4.4.6). If $u \in V$, the cyclic $k[x]$-submodule of V_ϕ generated by u is therefore equal to

$$k[\phi]u = \{p(\phi)u \mid p \in k[x]\} = k\phi^{n-1}u + \cdots + k\phi u + ku.$$

Since ϕ maps this subspace to itself, we say $k[\phi]u$ is *ϕ-invariant*. If u is nonzero, the $k[x]$-module homomorphism $k[x] \to k[\phi]u$ is onto. The kernel is a principal ideal $I_u = (q)$, and we have

$$k[\phi]u \cong k[x]/(q).$$

The polynomial q is called the *order of u*. Since u is nonzero and $k[\phi]u$ is finite dimensional over k, by Lemma 4.4.3 we know q is a monic polynomial of positive degree. In fact, q is the polynomial of minimal degree such that $q(\phi)u = 0$. By Exercise 4.5.45, q is a divisor of the minimal polynomial f of ϕ. Because the dimension of the k-vector space $k[\phi]u$ is equal to the degree of q, we see that q is the minimal polynomial of the restriction of ϕ to the ϕ-invariant subspace $k[\phi]u$.

For reference we list in Proposition 4.5.27 the fundamental results on cyclic $k[\phi]$-modules derived in the previous paragraphs.

PROPOSITION 4.5.27. *Let k be a field, V a k-vector space of dimension n, and ϕ a nonzero linear transformation in $\mathrm{Hom}_k(V,V)$. Let V_ϕ be the $k[x]$-module structure on V induced by the ring homomorphism $k[x] \to \mathrm{Hom}_k(V,V)$ which maps x to ϕ. If V_ϕ is a cyclic $k[x]$-module with generator u, then the following are true.*

(1) The set $B = \{u, \phi u, \phi^2 u, \ldots, \phi^{n-1}u\}$ is a k-basis for V.
(2) As $k[x]$-modules, $V_\phi \cong k[x]/(f)$.
(3) If $\min.\mathrm{poly}_k(\phi) = f$, then $\deg f = n$ and f is the monic polynomial of minimal degree such that $f(\phi)u = 0$.

PROOF. See the paragraphs immediately preceding the proposition. □

EXAMPLE 4.5.28. The ring of matrices $M_n(k)$ is a k-algebra where we identify k with the set of scalar matrices. The center of the ring of matrices is k. By Proposition 4.5.5, $\dim_k(M_n(k)) = n^2$. Since $M_n(k)$ is finite dimensional over k, every matrix $A \in M_n(k)$ has a minimal polynomial $\text{min.poly}_k(A)$ (see Theorem 4.4.6). The evaluation homomorphism $\theta : k[x] \to M_n(R)$ which is defined by $x \mapsto A$ maps $k[x]$ onto the commutative subring $k[A]$ of $M_n(R)$. The kernel of θ is the principal ideal generated by $f = \text{min.poly}_k(A)$.

$$\begin{array}{ccccc} k[x] & \xrightarrow{\theta} & k[A] & \xrightarrow{\subseteq} & M_n(R) \\ & \searrow^{\eta} & \uparrow{\scriptstyle \cong} & & \\ & & k[x]/(f) & & \end{array}$$

EXAMPLE 4.5.29. Let k be a field, $n \geq 2$, and $A = M_n(k)$ the ring of n-by-n matrices over k. Let e_{st} be the elementary matrix with 1 in position (s,t) and 0 elsewhere (see Definition 4.5.2). Notice that

$$e_{st}e_{uv} = \begin{cases} e_{sv} & \text{if } t = u, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $e_{st}e_{st} = 0$ if $s \neq t$ and $e_{ss}e_{ss} = e_{ss}$. From this it follows that

$$\text{min.poly}_k(e_{st}) = \begin{cases} x^2 - x & \text{if } s = t, \\ x^2 & \text{if } s \neq t. \end{cases}$$

In both cases we see that the minimal polynomial of e_{st} is not irreducible.

$$k[e_{st}] \cong \begin{cases} k[x]/x^2 - x & \text{if } s = t, \\ k[x]/x^2 & \text{if } s \neq t. \end{cases}$$

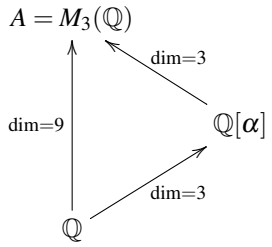
Therefore, $k[e_{st}]$ is not a field.

EXAMPLE 4.5.30. Let k be a field, $a \in k$, $A = M_3(k)$ the ring of 3-by-3 matrices over k, and $\alpha = \begin{bmatrix} 0 & 0 & a \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$. Notice that $\alpha^2 = \begin{bmatrix} 0 & a & 0 \\ 0 & 0 & a \\ 1 & 0 & 0 \end{bmatrix}$ and $\alpha^3 = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{bmatrix} = aI_3$. Therefore, α^3 is in k. Let $p(x) = x^3 - a$. Then $p(\alpha) = 0$. Let $f(x) = \text{min.poly}_k(\alpha)$. Then $f(x)$ divides $p(x)$. To show that $f(x)$ is equal to $p(x)$, it suffices to show $f(x)$ has degree greater than 2. First, since α is not a diagonal matrix we know $f(x)$ has degree greater than 1. For contradiction's sake, suppose $f(x) = x^2 + bx + c$ for some $b, c \in k$. Then $\alpha^2 + b\alpha \in k$. But

$$\alpha^2 + b\alpha = \begin{bmatrix} 0 & a & 0 \\ 0 & 0 & a \\ 1 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & ab \\ b & 0 & 0 \\ 0 & b & 0 \end{bmatrix} = \begin{bmatrix} 0 & a & ab \\ b & 0 & a \\ 1 & b & 0 \end{bmatrix}$$

is not a diagonal matrix. This contradiction implies $f(x)$ has degree greater than 2, hence $\text{min.poly}_k(\alpha) = x^3 - a$. This example is a special case of Exercise 4.6.21. The matrix α is called the companion matrix of the polynomial $x^3 - a$. Notice that $k[\alpha] \cong k[x]/(x^3 - a)$ is a field if and only if $x^3 - a$ is irreducible in $k[x]$. For instance, if $k = \mathbb{Q}$, and $a = 8$, then $x^3 - 8 = (x-2)(x^2 + 2x + 4)$ is not irreducible, hence $\mathbb{Q}[\alpha]$ is not a field. On the other hand, if $k = \mathbb{Q}$ and $a = 10$, then α is a root of $x^3 - 10$ in $M_3(\mathbb{Q})$, $\mathbb{Q}[\alpha]$ is an extension field

of k inside of A, and there is a lattice of subrings



where an arrow denotes set containment. Using the fact that $\mathbb{Q}[\alpha]$ is a subring of A we can view A as a vector space over $\mathbb{Q}[\alpha]$. We have $9 = (A : \mathbb{Q}) = (\mathbb{Q}[\alpha] : \mathbb{Q})(\mathbb{Q}[\alpha] : \mathbb{Q}) = 3 \cdot 3$. Notice that $\mathbb{Q}[\alpha]$ is not contained in the center of A, hence A is not an algebra over $\mathbb{Q}[\alpha]$.

5.4.2. *Rational Canonical Form.*

THEOREM 4.5.31. *If V is a finite dimensional vector space over the field k, and ϕ is a nonzero linear transformation in* $\mathrm{Hom}_k(V,V)$*, then there is a basis $\{u_1, u_2, \ldots, u_r\}$ for the $k[\phi]$-module V such that the following are true.*

(1) The $k[\phi]$-module V is equal to the internal direct sum $U_1 \oplus U_2 \oplus \cdots \oplus U_r$ where $U_i = k[\phi]u_i$ is the cyclic submodule of V spanned by u_i.
(2) $U_i \cong k[x]/(q_i)$ where q_i is the order of u_i and $q_1 \mid q_2 \mid \cdots \mid q_r$.
(3) U_i is a ϕ-invariant subspace of V and the minimal polynomial of $\phi|_{U_i}$ is q_i.
(4) The minimal polynomial of ϕ is q_r.
(5) The sequence of polynomials $(q_1, q_2, \ldots, q_r)$ is uniquely determined by ϕ.

The polynomials $q_1, \ldots, q_r$ are called the invariant factors *of ϕ.*

PROOF. Apply Theorem 4.3.7 to the finitely generated $k[x]$-module V. □

If V and ϕ are as in Theorem 4.5.31, then $V = U_1 \oplus \cdots \oplus U_r$ where each $\phi(U_i) \subseteq U_i$. Then each U_i is a k-subspace of V. We can pick a k-basis B_i for each subspace U_i and concatenate to get a basis $B = B_1 + \cdots + B_r$ for V. It is clear that the matrix of ϕ with respect to B is the block diagonal matrix

$$M(\phi, B) = \mathrm{diag}(M(\phi|_{U_1}, B_1), \ldots, M(\phi|_{U_r}, B_r))$$

where there are r blocks and block i is the matrix with respect to B_i of the restriction of ϕ to U_i.

Now we determine a canonical form for the matrix of ϕ. In other words, we try to find a basis B of V for which the matrix $M(\phi, B)$ is simplified. Based on the previous paragraph, we consider the case where $V = k[\phi]u$ is a cyclic module over the ring $k[\phi]$. We are in the context of Proposition 4.5.27. Suppose the minimal polynomial of ϕ is $\min.\mathrm{poly}_k(\phi) = p = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. The $k[x]$-module homomorphism $k[x] \to k[\phi]u$ defined by $1 \mapsto u$ is surjective and the kernel is the principal ideal $I_u = (p)$ generated by p. Therefore, as a $k[x]$-module, V is isomorphic to $k[x]/(p)$. Applying the division algorithm, we see that $1, x, x^2, \ldots, x^{n-1}$ is a k-basis for $k[x]/(p)$. Therefore, a k-basis for V is $B = \{u, \phi u, \phi^2 u, \ldots, \phi^{n-1}u\}$. Introduce the notation $x_i = \phi^{i-1}u$. The action of ϕ

on $B = \{x_1, x_2, \ldots, x_n\}$ determines the matrix $M(\phi, B)$. Computing, we get

$$\begin{aligned}
\phi x_1 &= \phi u = x_2 \\
\phi x_2 &= \phi\phi u = x_3 \\
&\vdots \\
\phi x_{n-1} &= \phi^{n-1} u = x_n \\
\phi x_n &= \phi^n u = -a_{n-1}\phi^{n-1}u - \cdots - a_1\phi^1 u - a_0 u = -a_0 x_1 - a_1 x_2 - \cdots - a_{n-1}x_n
\end{aligned}$$

so the matrix is

$$M(\phi, B) = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & 0 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & 0 & -a_1 \\ 0 & 1 & 0 & \ldots & 0 & 0 & -a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \\ 0 & 0 & 0 & \ldots & 0 & 0 & -a_{n-3} \\ 0 & 0 & 0 & \ldots & 1 & 0 & -a_{n-2} \\ 0 & 0 & 0 & \ldots & 0 & 1 & -a_{n-1} \end{bmatrix}. \tag{5.1}$$

We call (5.1) the *companion matrix* of the polynomial $p = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. If $p \in k[x]$ is a polynomial of degree $n \geq 1$, denote the companion matrix of p in $M_n(k)$ by $C(p)$. Conversely, by Exercise 4.5.46, the minimal polynomial of (5.1) is again $p = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$.

COROLLARY 4.5.32. *If V is a finite dimensional vector space over the field k, $\phi \in \mathrm{Hom}_k(V,V)$, and $q_1, q_2, \ldots, q_r$ are the invariant factors of ϕ, then there is a basis B for V such that the matrix of ϕ with respect to B is the block diagonal matrix*

$$M(\phi, B) = \mathrm{diag}\,(C(q_1), C(q_2), \ldots, C(q_r))$$

where block i is the companion matrix of q_i. The matrix $M(\phi, B)$ is called the rational canonical form *for ϕ.*

5.4.3. *Jordan Canonical Form.*

THEOREM 4.5.33. *If V is a finite dimensional vector space over the field k, and ϕ is a nonzero linear transformation in $\mathrm{Hom}_k(V,V)$, then there exist positive integers $s, \nu_1, \ldots, \nu_s$ and a basis $\{u_{ij} \mid 1 \leq i \leq s; 1 \leq j \leq \nu_i\}$ for the $k[\phi]$-module V such that the following are true.*

(1) The $k[\phi]$-module V is equal to the internal direct sum

$$V = \bigoplus_{i=1}^{s} \bigoplus_{j=1}^{\nu_i} U_{ij}$$

where $U_i = k[\phi]u_{ij}$ is the cyclic submodule of V spanned by u_{ij}.

(2) $U_{ij} \cong k[x]/(\pi_i^{e_{ij}})$ where

(a) $\pi_1, \ldots, \pi_s$ are distinct monic irreducible polynomials,

(b) the order of u_{ij} is $\pi_i^{e_{ij}}$, and

(c) $e_{i1} \geq e_{i2} \geq \cdots \geq e_{i\nu_i} \geq 1$.

(3) U_{ij} is a ϕ-invariant subspace of V and the minimal polynomial of $\phi|_{U_{ij}}$ is $\pi_i^{e_{ij}}$.

(4) The minimal polynomial of ϕ is

$$\mathrm{min.poly}_k(\phi) = \prod_{i=1}^{s} \pi_i^{e_{i1}}$$

(5) *The sequence of irreducible polynomials* $(\pi_1, \pi_2, \ldots, \pi_s)$ *and the positive integers* $\{e_{ij}\}$ *are uniquely determined by* ϕ.

The polynomials $\pi_i^{e_{ij}}$ *are called the* elementary divisors *of* ϕ.

PROOF. Apply Theorem 4.3.8 to the finitely generated $k[x]$-module V. □

Using the basis for V given by Theorem 4.5.33, we determine a canonical form for the matrix of ϕ. The minimal polynomial for ϕ restricted to U_{ij} is a power of the irreducible polynomial π_i. We assume each π_i is a linear polynomial, because the canonical form of ϕ in this case is particularly simplified. This case will occur if and only if the minimal polynomial of ϕ factors into a product of linear polynomials in $k[x]$. The k-bases for the individual ϕ-invariant subspaces U_{ij} can be concatenated for a basis of V. We now determine a canonical form for the matrix of ϕ under the following assumptions

(1) V is a cyclic $k[\phi]$-module spanned by u.
(2) $\min.\mathrm{poly}_k(\phi) = (x-b)^n$ is a power of a linear polynomial.

Notice that V is a cyclic $k[\phi]$-module, spanned by u. Since $k[\phi] = k[\phi - b]$, it follows that V is a cyclic $k[\phi - b]$-module, spanned by u. If $\theta : k[x] \to \mathrm{Hom}_k(V,V)$ is defined by $x \mapsto \phi$, then $\ker\theta$ is the principal ideal generated by $(x-b)^n$. If $\tau : k[x] \to \mathrm{Hom}_k(V,V)$ is defined by $x \mapsto \phi - b$, then the minimal polynomial of $\psi = \phi - b$ is the monic generator of $\ker\tau$, which is x^n. Therefore $B = \{u, \psi u, \psi^2 u, \ldots, \psi^{n-1}u\}$ is a k-basis for V. The matrix of $\psi = \phi - b$ with respect to the basis B is

$$M(\phi - b, B) = \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 & 0 \\ \vdots & & & \vdots & & & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 0 & \ldots & 1 & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 & 0 \end{bmatrix}$$

which is the companion matrix of the polynomial x^n. The matrix of ϕ with respect to the basis B is equal to $M(\phi, B) = M(\phi - b, B) + M(b, B)$. Therefore,

$$M(\phi, B) = \begin{bmatrix} b & 0 & 0 & \ldots & 0 & 0 & 0 \\ 1 & b & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & b & \ldots & 0 & 0 & 0 \\ \vdots & & & \vdots & & & \vdots \\ 0 & 0 & 0 & \ldots & b & 0 & 0 \\ 0 & 0 & 0 & \ldots & 1 & b & 0 \\ 0 & 0 & 0 & \ldots & 0 & 1 & b \end{bmatrix}. \tag{5.2}$$

We denote the n-by-n matrix (5.2) by $J_n(b)$ and refer to it as the basic *Jordan block* for the polynomial $(x-b)^n$.

COROLLARY 4.5.34. *Assume V is a finite dimensional vector space over the field k, $\phi \in \mathrm{Hom}_k(V,V)$, and that the minimal polynomial* $\min.\mathrm{poly}_k(\phi)$ *factors into a product of linear factors in $k[x]$. If $b_1, \ldots, b_s$ are the distinct roots of* $\min.\mathrm{poly}_k(\phi)$ *and $\{e_{ij}\}$ is the set of exponents of the elementary divisors of ϕ, then there is a basis B for V such that the matrix of ϕ with respect to B is the block diagonal matrix*

$$M(\phi, B) = \mathrm{diag}\left(J_{e_{11}}(b_1), J_{e_{12}}(b_1), \ldots, J_{e_{ij}}(b_i), \ldots\right)$$

where the block corresponding to the ordered pair (i,j) is the Jordan matrix of $(x-b_i)^{e_{ij}}$. The matrix $M(\phi,B)$ is called the Jordan canonical form *for ϕ and B is called a* Jordan basis.

5.4.4. *Canonical form of a matrix.* Let k be a field, and A a matrix in $M_n(k)$. With respect to the standard basis on $k^{(n)}$, left multiplication by A defines a linear transformation ℓ_A in $\mathrm{Hom}_k(k^{(n)},k^{(n)})$. The invariant factors, elementary divisors, rational canonical form, and the Jordan canonical form of A are defined to be the corresponding invariants of ℓ_A.

LEMMA 4.5.35. *Let V be a finite dimensional vector space over the field k. Let ϕ and ψ be linear transformations in $\mathrm{Hom}_k(V,V)$. The $k[x]$-modules V_ϕ and V_ψ are isomorphic if and only if there exists an invertible linear transformation ρ in $\mathrm{Hom}_k(V,V)$ such that $\phi=\rho^{-1}\psi\rho$.*

PROOF. Let $f:V_\phi\to V_\psi$ be an isomorphism of $k[x]$-modules. Then f is an isomorphism of k-vector spaces. That is, $f=\rho$ for some invertible element ρ in $\mathrm{Hom}_k(V,V)$. For each $u\in V$ we have $f(\phi u)=\psi f(u)$. Therefore, $\phi=\rho^{-1}\psi\rho$. Conversely, if $\phi=\rho^{-1}\psi\rho$, define $f:V_\phi\to V_\psi$ by $f(u)=\rho u$. For $i\geq 1$, we have $\rho\phi^i=\psi^i\rho$. Then $f(\phi^i u)=\rho\phi^i u=\psi^i\rho u=\psi^i f(u)$. The rest follows from the fact that ρ is k-linear. □

COROLLARY 4.5.36. *Let k be a field, and A and B two matrices in $M_n(k)$. The following are equivalent.*

(1) A and B are similar.
(2) A and B have the same invariant factors.
(3) A and B have the same rational canonical form.

PROOF. If A and B have the same invariant factors, say $q_1,q_2,\ldots,q_r$, then they are both similar to the block diagonal matrix $C=\mathrm{diag}\,(C(q_1),C(q_2),\ldots,C(q_r))$. The matrix C is in rational canonical form. The reader should verify that the invariant factors of C are $q_1,\ldots,q_r$. If A and B are similar, then by Proposition 4.5.8 and Lemma 4.5.35, the $k[x]$-modules that they induce on k^n are isomorphic. So they have the same invariant factors. □

EXAMPLE 4.5.37. Consider the matrix $A=\begin{bmatrix}1&1&1\\-1&-1&-1\\1&1&0\end{bmatrix}$ over the field $\mathbb{Q}$. Let $S=\{e_1,e_2,e_3\}$ be the standard basis for $V=\mathbb{Q}^{(3)}$. By Proposition 4.5.5, $A=M(\phi,S,S)$, where ϕ is the linear transformation in $\mathrm{Hom}_{\mathbb{Q}}(V,V)$ defined by multiplication by A from the left. Notice that $A^2=\begin{bmatrix}1&1&0\\-1&-1&0\\0&0&0\end{bmatrix}$, and $A^3=0$. Thus, A is nilpotent and the index of nilpotency is 3. This proves that $\mathrm{min.poly}(A)=x^3$. Since the minimal polynomial of A has only one root and is split, the rational canonical form of A is equal to the Jordan canonical form, which is $J_3(0)=\begin{bmatrix}0&0&0\\1&0&0\\0&1&0\end{bmatrix}$. Let $u_1=(1,0,0)^t$, $u_2=Au_1=(1,-1,1)^t$, and $u_3=Au_2=(1,-1,0)^t$. Then $B=\{u_1,u_2,u_3\}$ is a Jordan basis for ϕ. If $P=\begin{bmatrix}1&1&1\\0&-1&-1\\0&1&0\end{bmatrix}$

is the matrix with columns u_1, u_2, u_3, the reader should verify that $P^{-1} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & -1 \end{bmatrix}$ and $P^{-1}AP = J_3(0)$.

5.5. Reduced Row Echelon Form. In this section we show that any matrix over a field has a unique reduced row echelon form. This canonical form exists whether the matrix is square or not. Using gaussian elimination and elementary row operations, an algorithm which is not included in this book, the reduced row echelon form can be efficiently computed.

DEFINITION 4.5.38. Let k be a field and $R \in M_{mn}(k)$ an m-by-n matrix. We say R is in *reduced row echelon form*, if the following conditions are satisfied:

(1) Any row that consists only of zeros is below any nonzero row.
(2) The left-most nonzero entry of a row is equal to 1. We call this 1 a *leading* 1.
(3) The leading ones form a staggered, or echelon pattern from left to right and top to bottom. That is, if $i < j$ and rows i and j are nonzero, then the leading 1 in row i is to the left of the leading 1 in row j.
(4) Above and below any leading 1 are zeros.

LEMMA 4.5.39. *Let k be a field and $R \in M_{mn}(k)$ an m-by-n matrix in reduced row echelon form.*

(1) The rank of R is equal to the number of nonzero rows.
(2) The rank of R is equal to the number of leading ones.
(3) The nullity of R is equal to the number of columns that do not contain a leading 1.
(4) Let $R_1, \ldots, R_n$ be the columns of R. If R_j does not contain a leading 1*, then R_j is a unique linear combination of the columns in the set $\{R_1, \ldots, R_{j-1}\}$ that contain a leading one. In other words, there is a unique vector in the kernel of R of the form $(x_1, \ldots, x_{j-1}, 1, 0, \ldots, 0)$ such that for $1 \leq i < j$, $x_i = 0$ if R_i does not contain a leading* 1.

PROOF. The proof is left to the reader. □

PROPOSITION 4.5.40. *Let k be a field and $A \in M_{mn}(k)$.*

(1) There is an invertible matrix Q in $M_m(k)$ such that QA is in reduced row echelon form.
(2) The reduced row echelon form of A is unique in the sense that if Q_1 is another invertible matrix in $M_m(k)$ and Q_1A is in reduced row echelon form, then $QA = Q_1A$.

PROOF. (1): Let $X = \{A_1, A_2, \ldots, A_n\}$ be the columns of A. The column space of A is equal to the span of X in $k^{(m)}$. By Corollary 4.2.6 there exists a subset of X that is a basis for the column space of A. Let $U \subseteq X$ be a basis for the column space of A such that U is minimal with respect to the ordering on 2^X defined in Exercise 1.2.24. Then $U \subseteq X$ has the property that if $A_j \in X - U$, then A_j is a linear combination of $\{A_i \in U \mid i < j\}$. By Theorem 4.2.4, we can extend U to a basis for $k^{(m)}$. Call the resulting basis B. Let Q be the change of basis matrix. Then Q is an invertible matrix in $M_m(k)$. Let $QA = R$. We show that R is a matrix in reduced row echelon form. Let $\text{Rank}(A) = r$ and $M_U = (u_1, \ldots, u_r)$ the m-by-r matrix with columns the r vectors in U. Then QM_U is the m-by-r matrix equal

to the first r columns of the identity matrix I_m in $M_m(k)$. Therefore, the columns of A in U correspond to the standard basis vectors $e_1,\dots,e_r$ in R. The column space of R is spanned by $e_1,\dots,e_r$, hence rows $r+1,\dots,m$ of R are zeros. As mentioned above, if $A_j \in X-U$, then A_j is a linear combination of those columns of A that are in U and to the left of A_j. This says that every nonzero row of R has a leading one.

(2): Since Q is invertible, the kernel of ℓ_{QA} is equal to the kernel of ℓ_A. Suppose $Q_1A = R_1$ and $Q_2A = R_2$ are two reduced row echelon forms for A. For sake of contradiction, suppose there is a difference in the columns containing leading ones. Say there is a leading 1 in column i of R_1 but not in column i of R_2. Then this contradicts Lemma 4.5.39 (4) because a column containing a leading 1 is not linearly dependent on the columns to its left. The uniqueness of those columns that do not contain leading ones follows from Lemma 4.5.39 (4) and the fact that the kernels of ℓ_{R_1} and ℓ_{R_2} are equal. □

PROPOSITION 4.5.41. *Let k be a field, A a matrix in $M_{mn}(k)$, and Q an invertible matrix in $M_m(k)$ such that QA is in reduced row echelon form.*

(1) The columns of QA containing leading ones correspond to a set of columns of A that make up a basis for the column space of A.
(2) If A has rank r, then the $n-r$ vectors described in Lemma 4.5.39 (4) make up a basis for the kernel of A.

PROOF. The proof is left to the reader. □

EXAMPLE 4.5.42. Consider the matrix $A = \begin{bmatrix} 1 & 2 & -1 & 0 \\ 2 & 1 & 1 & 3 \\ 1 & -1 & 2 & 3 \end{bmatrix}$ over a field k, where we assume $\operatorname{char} k \neq 3$. Notice that $Q = \begin{bmatrix} -1/3 & 2/3 & 0 \\ 2/3 & -1/3 & 0 \\ 1 & -1 & 1 \end{bmatrix}$ is invertible and the inverse is $Q^{-1} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix}$. Multiplying, $QA = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ is in reduced row echelon form. The rank of A is 2, the nullity of A is 2. The first two columns of A make up a basis for the column space of A. From Lemma 4.5.39 (4), we obtain a basis for the kernel of A by writing columns 3 and 4 of QA as linear combinations of columns 1 and 2:

$$\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

A basis for the kernel of A is $(-1,1,1,0)^t, (-2,1,0,1)^t$.

5.6. A System of Linear Equations. Let k be a field. Consider a system of m linear equations in n variables over k:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad\quad \vdots \qquad\qquad \vdots \qquad\quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{5.3}$$

Then the matrix of coefficients $A = (a_{ij})$ is in $M_{mn}(k)$ and the vector $b = (b_1,\dots,b_m)^t$ on the right-hand side is in $k^{(m)}$. If $x = (x_1,\dots,x_n)^t$, then (5.3) can be expressed in matrix form: $Ax = b$. With respect to the standard bases on $k^{(n)}$ and $k^{(m)}$, left multiplication by A

defines a linear transformation $T = \ell_A$ in $\mathrm{Hom}_k(k^{(n)}, k^{(m)})$. The image of T is the column space of A. The rank of A is the dimension of the column space of T. The nullity of A is the dimension of the kernel of T.

PROPOSITION 4.5.43. *In the above context,*

(1) If b is in the image of T, then the system of linear equations (5.3) *has a solution. Let $c = (c_1, \ldots, c_n)^t$ be a particular solution. Then the general solution to* (5.3) *is $x = c + z$, where $z = (z_1, \ldots, z_n)^t$ represents a typical element in the kernel of T. The nullity of T is equal to the number of degrees of freedom in the solution. The solution x is unique if and only if the nullity of T is zero. If the nullity of T is positive, then we say the system of equations is* underdetermined.

(2) If b is not in the image of T, then there is no solution to (5.3)*. In this case, we say the system of equations is* overdetermined.

PROOF. The proof is left to the reader. □

EXAMPLE 4.5.44. This is a continuation of Example 4.5.42. Consider the system of 3 linear equations in 4 variables:

$$\begin{aligned} x_1 + 2x_2 - x_3 &= 2 \\ 2x_1 + x_2 + x_3 + 3x_4 &= 7 \\ x_1 - x_2 + 2x_3 + 3x_4 &= 5 \end{aligned}$$

Then the matrix of coefficients is $A = \begin{bmatrix} 1 & 2 & -1 & 0 \\ 2 & 1 & 1 & 3 \\ 1 & -1 & 2 & 3 \end{bmatrix}$ and the right-hand side vector is $b = (2,7,5)^t$. From Example 4.5.42, the reduced row echelon form of A is obtained by multiplying by $Q = \begin{bmatrix} -1/3 & 2/3 & 0 \\ 2/3 & -1/3 & 0 \\ 1 & -1 & 1 \end{bmatrix}$. Let $x = (x_1, x_2, x_3, x_4)^t$. A basis for the kernel of A is $(-1,1,1,0)^t, (-2,1,0,1)^t$. Multiply both sides of the matrix equation $Ax = b$ by Q:

$$QAx = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix}$$

Then the general solution is:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \\ 0 \\ 0 \end{bmatrix} + a \begin{bmatrix} -1 \\ 1 \\ 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

where a and b represent arbitrary elements of k.

5.7. Exercises.

EXERCISE 4.5.45. Let k be a field, V a finite dimensional k-vector space, u a nonzero vector in V, and $\phi \in \mathrm{Hom}_k(V,V)$. Let $f \in k[x]$ be the monic polynomial of minimal degree such that $f(\phi)u = 0$. Prove that f divides $\min.\mathrm{poly}_k(\phi)$.

EXERCISE 4.5.46. Let k be a field, V a k-vector space of dimension n, and $\phi \in \mathrm{Hom}_k(V,V)$. Suppose $B = \{x_1, \ldots, x_n\}$ is a k-basis for V and $\{a_0, \ldots, a_{n-1}\} \subseteq k$ such that $\phi x_1 = x_2$, $\phi x_2 = x_3$, …, $\phi x_{n-1} = x_n$, and $\phi x_n = -a_0 x_1 - a_1 x_2 - \cdots - a_{n-1} x_n$. Prove:

(1) $V_\phi = k[\phi]x_1$. In other words, V_ϕ is a cyclic $k[\phi]$-module and is generated by x_1.
(2) $\min.\text{poly}_k(\phi) = x^n + a_{n-1}x_{n-1} + \cdots + a_1x + a_0$.

EXERCISE 4.5.47. Assume A is an n-by-n matrix over the field $\mathbb{Q}$ such that the minimum polynomial of A in $\mathbb{Q}[x]$ is equal to $(x^2+1)(x+2)$. If $n = 7$, exhibit all possible rational canonical forms for A.

EXERCISE 4.5.48. Let k be a field. Let q and ℓ be monic polynomials in $k[x]$, where q is an irreducible quadratic and ℓ is linear. If A is a 7-by-7 matrix over k such that the minimum polynomial of A in $k[x]$ is $q\ell$, exhibit all possible rational canonical forms for A.

EXERCISE 4.5.49. Let k be a field. Let q and ℓ be monic polynomials in $k[x]$, where q is an irreducible quadratic and ℓ is linear. Let A be a 6-by-6 matrix over k. Exhibit all possible rational canonical forms for A, if the minimum polynomial of A in $k[x]$ is $q^2\ell$. Do the same if the minimum polynomial of A in $k[x]$ is $\ell^2 q$.

EXERCISE 4.5.50. Let k be a field. Let q and t be irreducible monic polynomials in $k[x]$, where $\deg q = 2$ and $\deg t = 3$. Let A be a 15-by-15 matrix over k. Exhibit all possible rational canonical forms for A, if the minimum polynomial of A in $k[x]$ is q^2t^2. Do the same if the minimum polynomial of A in $k[x]$ is q^3t.

EXERCISE 4.5.51. Let k be a field. Let q_1, q_2 and ℓ be distinct irreducible monic polynomials in $k[x]$, where q_1 and q_2 are quadratics and ℓ is linear. Let A be a 10-by-10 matrix over k. Exhibit all possible rational canonical forms for A, if the minimum polynomial of A in $k[x]$ is $\ell q_1^2 q_2$.

EXERCISE 4.5.52. Let k be a field. Let ℓ_1, ℓ_2 be distinct monic polynomials in $k[x]$, where $\deg \ell_1 = \deg \ell_2 = 1$. Let A be an 8-by-8 matrix over k. Exhibit all possible rational canonical forms for A, if the minimum polynomial of A in $k[x]$ is $\ell_1^2\ell_2^3$.

EXERCISE 4.5.53. Let F/k be an extension of fields. Prove the following.

(1) Let $A \in M_n(k)$. Then A is invertible in $M_n(k)$ if and only if A is invertible in $M_n(F)$. (Hint: Theorem 4.4.11.)
(2) If X is a basis for $k^{(n)}$ over the field k, then X is a basis for $F^{(n)}$ over the field F.
(3) If $A \in M_n(k)$, then the invariant factors of A in $k[x]$ are the same as the invariant factors of A in $F[x]$.
(4) Let $A, B \in M_n(k)$. Then A is similar to B in $M_n(k)$ if and only if A is similar to B in $M_n(F)$.

EXERCISE 4.5.54. Let k be a field and $b \in k$. Let $B \in M_n(k)$ be the Jordan block corresponding to $(x-b)^n$. That is, B is the matrix which has main diagonal entries all equal to b, first lower subdiagonal entries all equal to 1 and 0 elsewhere. Prove that the transpose of B is similar to B. For a continuation of this exercise, see Exercise 5.2.24.

6. The Determinant

Throughout this section, R is a commutative ring and n is a fixed positive integer. We prove that the determinant function $\det : M_n(R) \to R$ exists and is the unique alternating multilinear form (on the columns) such that if I_n is the identity matrix, then $\det(I_n) = 1$.

Let $J = \{1, \ldots, n\}$ and $J^n = J \times \cdots \times J$ (n times). We view the symmetric group S_n as the subset of J^n consisting of n-tuples $\vec{j} = (j_1, \ldots, j_n)$ that are permutations of J. The sign of a permutation $\sigma \in S_n$ is denoted $\text{sign}(\sigma)$.

DEFINITION 4.6.1. Let R be a commutative ring, $n \geq 1$, and $(R^n)^n = \bigoplus_{i=1}^n R^n$. Consider a function $f : (R^n)^n \to R$. We say that f is a *multilinear form* if for each i,

$$f(x_1,\ldots,x_{i-1},\alpha u+\beta v,x_{i+1},\ldots,x_n) = \\ \alpha f(x_1,\ldots,x_{i-1},u,x_{i+1},\ldots,x_n)+\beta f(x_1,\ldots,x_{i-1},v,x_{i+1},\ldots,x_n).$$

We say that f is an *alternating form* if $f(x_1,,\ldots,x_n) = 0$ whenever $x_i = x_j$ for some pair $i \neq j$.

LEMMA 4.6.2. *If $f : (R^n)^n \to R$ is an alternating multilinear form and $\sigma \in S_n$ is a permutation on the set $\{1,\ldots,n\}$, then*

$$f(x_{\sigma 1},\ldots,x_{\sigma n}) = \operatorname{sign}(\sigma) f(x_1,,\ldots,x_n).$$

We say that f is skew symmetric.

PROOF. Because σ factors into a product of transpositions, it is enough to show that acting on the variables by a transposition changes the sign of f. For simplicity's sake, assume $\sigma = (i,j) = (1,2)$. Look at

$$\begin{aligned} 0 &= f(x_1+x_2,x_1+x_2,x_3,\ldots,x_n) \\ &= f(x_1,x_1,x_3,\ldots,x_n)+f(x_1,x_2,x_3,\ldots,x_n)+ \\ &\quad f(x_2,x_1,x_3,\ldots,x_n)+f(x_2,x_2,x_3,\ldots,x_n) \\ &= f(x_1,x_2,x_3,\ldots,x_n)+f(x_2,x_1,x_3,\ldots,x_n). \end{aligned}$$

This shows $f(x_1,x_2,x_3,\ldots,x_n) = -f(x_2,x_1,x_3,\ldots,x_n)$. □

LEMMA 4.6.3. *If R is a commutative ring and $r \in R$, there is a unique alternating multilinear form $f : (R^n)^n \to R$ such that $f(e_1,\ldots,e_n) = r$, where $(e_1,\ldots,e_n)$ is the standard basis for R^n.*

PROOF. (Uniqueness) Given $(x_1,\ldots,x_n) \in (R^n)^n$, for each i we can write $x_i = a_{1i}e_1 + \cdots + a_{ni}e_n$. Since f is multilinear,

$$\begin{aligned} f(x_1,\ldots,x_n) &= f\left(\sum_{j\in J} a_{j1}e_j,\ldots,\sum_{j\in J} a_{jn}e_j\right) \\ &= \sum_{j_1\in J}\left(a_{j_1 1} f\left(e_{j_1},\sum_{j\in J} a_{j2}e_j,\ldots,\sum_{j\in J} a_{jn}e_j\right)\right) \\ &= \sum_{j_1\in J}\sum_{j_2\in J}\left(a_{j_1 1}a_{j_2 2} f\left(e_{j_1},e_{j_2},\ldots,\sum_{j\in J} a_{jn}e_j\right)\right) \\ &\vdots \\ &= \sum_{(j_1,\ldots,j_n)\in J^n} a_{j_1 1}\cdots a_{j_n n} f\left(e_{j_1},\ldots,e_{j_n}\right). \end{aligned}$$

Since f is alternating, if $\vec{j} = (j_1,\ldots,j_n) \in J^n$ is not a permutation, then $f(e_{j_1},\ldots,e_{j_n}) = 0$. We can restrict the latest summation to $\vec{j} \in S_n$. In this case, since f is skew symmetric, $f(e_{j_1},\ldots,e_{j_n}) = \operatorname{sign}(j) f(e_1,\ldots,e_n) = \operatorname{sign}(j) r$. This proves that

$$f(x_1,\ldots,x_n) = r\sum_{\vec{j}\in S_n} \operatorname{sign}(\vec{j}) a_{j_1 1}\cdots a_{j_n n} \tag{6.1}$$

is completely determined by r and $(x_1,\ldots,x_n)$.

(Existence) The formula in (6.1) defines a function $f : (R^n)^n \to R$. Notice that

$$f(e_1, \ldots, e_n) = r$$

since only for $\vec{j} = (1, 2, \ldots, n)$ is the product formula in the summation (6.1) nonzero. We need to prove f is an alternating multilinear form. Let $\alpha, \beta \in R$, $u, v \in R^n$. Write $u = \sum u_i e_i$ and $v = \sum v_i e_i$. Set $a_{ik} = \alpha u_i + \beta v_i$, so that $x_k = \sum a_{ik} e_i = \sum (\alpha u_i + \beta v_i) e_i = \alpha u + \beta v$. Then

$$\begin{aligned}
f(x_1, \ldots, \alpha u + \beta v, \ldots, x_n) &= r \sum_{\vec{j} \in S_n} \operatorname{sign}(j) a_{j_1 1} \cdots a_{j_k k} \cdots a_{j_n n} \\
&= r \sum_{\vec{j} \in S_n} \operatorname{sign}(j) a_{j_1 1} \cdots (\alpha u_{j_k} + \beta v_{j_k}) \cdots a_{j_n n} \\
&= r\alpha \sum_{\vec{j} \in S_n} \operatorname{sign}(j) a_{j_1 1} \cdots u_{j_k} \cdots a_{j_n n} + \\
&\qquad r\beta \sum_{\vec{j} \in S_n} \operatorname{sign}(j) a_{j_1 1} \cdots v_{j_k} \cdots a_{j_n n} \\
&= \alpha f(x_1, \ldots, u, \ldots, x_n) + \beta f(x_1, \ldots, v, \ldots, x_n)
\end{aligned}$$

shows f is multilinear.

Now we show f is alternating. Suppose $i < j$ and let τ be the transposition that switches i and j. The alternating group A_n has index 2 in S_n, so every odd permutation is of the form $\sigma\tau$ for some $\sigma \in A_n$. Assume $x_i = x_j$ and show $f(x_1, \ldots, x_n) = 0$. For all k we have $a_{ki} = a_{kj}$. Also, if $\sigma \in A_n$ then $\sigma\tau(i) = \sigma(j)$ and $\sigma\tau(j) = \sigma(i)$.

$$\begin{aligned}
f(x_1, \ldots, x_n) &= r \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n} \\
&= r \sum_{\sigma \in A_n} \left(a_{\sigma(1)1} \cdots a_{\sigma(n)n} - a_{\sigma\tau(1)1} \cdots a_{\sigma\tau(n)n} \right) \\
&= r \sum_{\sigma \in A_n} \left(a_{\sigma(1)1} \cdots a_{\sigma(n)n} - a_{\sigma\tau(1)1} \cdots a_{\sigma\tau(i)i} \cdots a_{\sigma\tau(j)j} \cdots a_{\sigma\tau(n)n} \right) \\
&= r \sum_{\sigma \in A_n} \left(a_{\sigma(1)1} \cdots a_{\sigma(n)n} - a_{\sigma(1)1} \cdots a_{\sigma(j)i} \cdots a_{\sigma(i)j} \cdots a_{\sigma(n)n} \right) \\
&= r \sum_{\sigma \in A_n} \left(a_{\sigma(1)1} \cdots a_{\sigma(n)n} - a_{\sigma(1)1} \cdots a_{\sigma(j)j} \cdots a_{\sigma(i)i} \cdots a_{\sigma(n)n} \right) \\
&= 0.
\end{aligned}$$

□

DEFINITION 4.6.4. By viewing the columns of a matrix in $M_n(R)$ as vectors in R^n, we identify $M_n(R)$ with $(R^n)^n$. The *determinant* is the unique alternating multilinear form $\det : M_n(R) \to R$ such that $\det(I_n) = 1$. By Lemma 4.6.3,

$$\det(a_{ij}) = \sum_{\vec{j} \in S_n} \operatorname{sign}(j) a_{j_1,1} \cdots a_{j_n,n}.$$

LEMMA 4.6.5. *Let $A, B \in M_n(R)$.*

(1) $\det(AB) = \det(A)\det(B)$.
(2) A is invertible if and only if $\det(A)$ *is a unit in R.*
(3) If A and B are similar, then $\det(A) = \det(B)$.
(4) $\det(A) = \det(A^T)$.
(5) The determinant is an alternating multilinear form on the rows of matrices in $M_n(R)$.

PROOF. (1): Fix A. Taking $r = \det(A)$ in (6.1) defines an alternating multilinear form $g : M_n(R) \to R$, where $g(C) = \det(A)\det(C)$. Define another function $f : M_n(R) \to R$ by $f(C) = \det(AC)$. Since $f(I_n) = \det(A)$, by Lemma 4.6.3, it is enough to prove that f is alternating and multilinear. Assume $u, v \in R^n$ and $C = (c_1, \ldots, c_n) \in M_n(R)$. Then

$$\begin{aligned} f(c_1, \ldots, \alpha u + \beta v, \ldots, c_n) &= \det(A(c_1, \ldots, \alpha u + \beta v, \ldots, c_n)) \\ &= \det(Ac_1, \ldots, \alpha Au + \beta Av, \ldots, Ac_n) \\ &= \alpha \det(Ac_1, \ldots, Au, \ldots, Ac_n) + \beta \det(Ac_1, \ldots, Av, \ldots, Ac_n) \\ &= \alpha f(c_1, \ldots, u, \ldots, c_n) + \beta f(c_1, \ldots, v, \ldots, c_n). \end{aligned}$$

If two columns of C are equal, then two columns of AC are equal, so f is alternating.

(2): If $AB = I_n$, then $\det(A)\det(B) = 1$. The converse follows from Lemma 4.6.9 because in this case $A^{-1} = \det(A)^{-1}A^a$.

(3): If $A = X^{-1}BX$, then

$$\begin{aligned} \det(A) &= \det(X^{-1})\det(B)\det(X) \\ &= \det(B)\det(X^{-1})\det(X) \\ &= \det(B)\det(X^{-1}X) \\ &= \det(B). \end{aligned}$$

(4): Since R is commutative, for every $\sigma \in S_n$ we have

$$a_{\sigma(1),1} \cdots a_{\sigma(n),n} = a_{1,\sigma^{-1}(1)} \cdots a_{n,\sigma^{-1}(n)}.$$

This together with the fact that $\operatorname{sign}(\sigma) = \operatorname{sign}(\sigma^{-1})$ lead to

$$\begin{aligned} \det(A) &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n} \\ &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1,\sigma^{-1}(1)} \cdots a_{n,\sigma^{-1}(n)} \\ &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} \\ &= \det(A^T). \end{aligned}$$

(5): Follows from (4). □

DEFINITION 4.6.6. For $A \in M_n(R)$, let A_{ij} be the matrix in $M_{n-1}(R)$ obtained by deleting row i and column j from A. Then $\det(A_{ij})$ is called the *minor* of A in position (i, j) and $(-1)^{i+j}\det(A_{ij})$ is called the *cofactor* of A in position (i, j).

LEMMA 4.6.7. *If A is a matrix in $M_n(R)$, then the following are true.*

(1) For each row i, $\det(A) = \sum_{j=1}^n a_{ij}(-1)^{i+j}\det(A_{ij})$, and
(2) For each column j, $\det(A) = \sum_{i=1}^n a_{ij}(-1)^{i+j}\det(A_{ij})$.

PROOF. We prove that the determinant can be computed by cofactor expansion of row i. The statement about column expansion follows from Lemma 4.6.5 (4). Define a function $f : M_n(R) \to R$ by the formula $f(A) = \sum_{j=1}^n a_{ij}(-1)^{i+j}\det(A_{ij})$. The reader should verify that $f(I_n) = 1$. By Lemma 4.6.3 it is enough to show that f is alternating and multilinear.

Assume the columns of A are $(A_1, \ldots, A_n)$ and assume $A_k = A_\ell$ and $k < \ell$. Therefore $a_{ik} = a_{i\ell}$. If $j \neq k$ and $j \neq \ell$, then A_{ij} has two columns that are equal, so $\det(A_{ij}) = 0$. The

formula for f reduces to

$$\begin{aligned} f(A) &= a_{ik}(-1)^{i+k}\det(A_{ik}) + a_{i\ell}(-1)^{i+\ell}\det(A_{i\ell}) \\ &= a_{ik}(-1)^{i+k}\det(A_{ik}) + a_{ik}(-1)^{i+\ell}\det(A_{i\ell}) \\ &= a_{ik}\left((-1)^{i+k}\det(A_{ik}) + (-1)^{i+\ell}\det(A_{i\ell})\right). \end{aligned}$$

But A_{ik} is obtained from $A_{i\ell}$ by permuting the columns. In fact, $\ell-k-1$ transpositions are sufficient. Since the determinant form is skew symmetric, $\det(A_{ik}) = (-1)^{\ell-k-1}\det(A_{i\ell})$. The reader should verify that $(-1)^{i+k} + (-1)^{i+\ell}(-1)^{\ell-k-1} = 0$, hence

$$\begin{aligned} f(A) &= a_{ik}\left((-1)^{i+k}\det(A_{ik}) + (-1)^{i+\ell}\det(A_{i\ell})\right) \\ &= a_{ik}\left((-1)^{i+k}\det(A_{ik}) + (-1)^{i+\ell}(-1)^{\ell-k-1}\det(A_{ik})\right) \\ &= a_{ik}\det(A_{ik})\left((-1)^{i+k} + (-1)^{i+\ell}(-1)^{\ell-k-1}\right) \\ &= 0 \end{aligned}$$

which proves f is alternating.

Assume the columns of A are $(A_1,\ldots,A_n)$ where $A_k = \alpha u + \beta v$ for some $u, v \in R^n$. Let $B = (b_{ij})$ be the matrix obtained by replacing column k of A with the vector u. Let $C = (c_{ij})$ be the matrix obtained by replacing column k of A with the vector v. We show that $f(A) = \alpha f(B) + \beta f(C)$. Because they differ only in column k, we have $A_{ik} = B_{ik} = C_{ik}$. If $j \neq k$, then the determinant is multilinear, so $\det(A_{ij}) = \alpha\det(B_{ij}) + \beta\det(C_{ij})$. Therefore

$$\begin{aligned} f(A) &= \sum_{j=1}^{n} a_{ij}(-1)^{i+j}\det(A_{ij}) \\ &= \sum_{j\neq k} a_{ij}(-1)^{i+j}\left(\alpha\det(B_{ij}) + \beta\det(C_{ij})\right) + (\alpha b_{ik} + \beta c_{ik})(-1)^{i+k}\det(A_{ik}) \\ &= \alpha\sum_{j=1}^{n} b_{ij}(-1)^{i+j}\det(B_{ij}) + \beta\sum_{j=1}^{n} c_{ij}(-1)^{i+j}\det(C_{ij}) \\ &= \alpha f(B) + \beta f(C) \end{aligned}$$

□

DEFINITION 4.6.8. Let $A \in M_n(R)$. The *adjoint* of A, denoted A^a, is the transpose of the matrix of cofactors of A. Therefore, $A^a = \left((-1)^{i+j}\det(A_{ji})\right)$.

LEMMA 4.6.9. $A^aA = AA^a = \det(A)I_n$.

PROOF. Assume $i \neq j$. Let B be the matrix which is equal to A with column i replaced with a copy of column j. Compute $\det(B) = 0$ by column expansion down column i. Use the facts that $B_{ki} = A_{ki}$ and $b_{ki} = b_{kj} = a_{kj}$ for each k.

$$\begin{aligned} 0 &= \sum_{k=1}^{n} b_{ki}(-1)^{i+k}\det(B_{ki}) \\ &= \sum_{k=1}^{n} a_{kj}(-1)^{i+k}\det(A_{ki}) \end{aligned}$$

Let $A^aA = (c_{ij})$. Then

$$c_{ij} = \sum_{k=1}^{n}(-1)^{i+k}\det(A_{ki})a_{kj} = \begin{cases} \det(A) & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

□

DEFINITION 4.6.10. By Lemma 4.6.5, the determinant function is constant on similarity classes. If M is a finitely generated free R-module and $\phi \in \mathrm{Hom}_R(M,M)$, then the determinant of ϕ is defined to be the determinant of the matrix of ϕ with respect to any basis of M.

6.1. The Characteristic Polynomial.

DEFINITION 4.6.11. Let k be a field, V a finite dimensional k-vector space, and $T \in \mathrm{Hom}_k(V,V)$. If T is not invertible, then we say T is *singular*.

THEOREM 4.6.12. *Let k be a field, V a finite dimensional k-vector space, and $T \in \mathrm{Hom}_k(V,V)$. The following are true.*

(1) $T \neq 0$ *if and only if there exists* $v \in V$ *such that* $T(v) \neq 0$.
(2) $\min.\mathrm{poly}_k(T)$ *has degree less than or equal to* n^2.
(3) $k[T]$ *is a commutative k-subalgebra of* $\mathrm{Hom}_k(V,V)$.
(4) The following are equivalent.
  (a) T *is singular.*
  (b) The constant term of $\min.\mathrm{poly}_k(T)$ *is zero.*
  (c) There exists $S \in \mathrm{Hom}_k(V,V)$ *such that* $S \neq 0$ *and* $TS = ST = 0$.
  (d) There exists $v \in V - (0)$ *such that* $T(v) = 0$.
(5) The following are equivalent.
  (a) T is invertible.
  (b) $\mathrm{Rank}(T) = \dim_k(V)$.
  (c) $\mathrm{Nullity}(T) = 0$.

PROOF. For the proof, apply Proposition 4.5.5, Theorems 4.4.6 and 4.4.11, Corollaries 4.4.8 and 4.4.10, and Exercise 4.2.11. □

DEFINITION 4.6.13. Let R be a commutative ring and $M \in M_n(R)$. If x is an indeterminate, then we can view M as a matrix in $M_n(R[x])$. The *characteristic polynomial* of M is $\mathrm{char.poly}_R(M) = \det(xI_n - M)$, which is a polynomial in $R[x]$. Computing the determinant using row expansion (Lemma 4.6.7) along row one, it is easy to see that $\mathrm{char.poly}_R(M)$ is monic and has degree n. The characteristic polynomial is constant on similarity classes, by Exercise 4.6.22. If P is a finitely generated free R-module and $\phi \in \mathrm{Hom}_R(P,P)$, then the characteristic polynomial of ϕ is defined to be the characteristic polynomial of the matrix of ϕ with respect to any basis of P.

THEOREM 4.6.14. *Let k be a field and V a finite dimensional vector space over k. Let $\phi \in \mathrm{Hom}_k(V,V)$. As in Theorem 4.5.31, let $q_1, q_2, \ldots, q_r$ be the invariant factors of ϕ.*

(1) $\mathrm{char.poly}_k(\phi) = q_1 q_2 \cdots q_r$.
(2) (Cayley-Hamilton) If $p(x) = \mathrm{char.poly}_k(\phi)$, *then* $p(\phi) = 0$. *The minimal polynomial of ϕ divides the characteristic polynomial of ϕ. That is,* $\min.\mathrm{poly}_k(\phi) \mid \mathrm{char.poly}_k(\phi)$.
(3) If $f \in k[x]$ is irreducible, then $f \mid \mathrm{char.poly}_k(\phi)$ *if and only if* $f \mid \min.\mathrm{poly}_k(\phi)$. *The roots of* $\min.\mathrm{poly}_k(\phi)$ *are precisely the roots of* $\mathrm{char.poly}_k(\phi)$.

PROOF. (1): By Corollary 4.5.32 there is a basis for V such that the matrix of ϕ is the block diagonal matrix $(C(q_1), C(q_2), \ldots, C(q_r))$, where $C(q_i)$ is the companion matrix for q_i. By Exercise 4.6.21, the characteristic polynomial of $C(q_i)$ is q_i. Apply Exercise 4.6.23 iteratively to show that $\mathrm{char.poly}_k(\phi) = q_1 q_2 \cdots q_r$.

(2): By Theorem 4.5.31, $\min.\text{poly}_k(\phi) = q_r$.

(3): By Theorem 4.5.31, $q_1 \mid q_2 \mid \cdots \mid q_r$. The irreducible factors of $\text{char.poly}_k(\phi)$ are equal to the irreducible factors of $\min.\text{poly}_k(\phi)$. □

DEFINITION 4.6.15. Let k be a field, V a finite dimensional k-vector space. If $\phi \in \text{Hom}_k(V,V)$ and $\lambda \in k$, then λ is called a *characteristic root* or *eigenvalue* of ϕ if $\phi - \lambda$ is singular. The set $U(\lambda) = \ker(\phi - \lambda) = \{x \in V \mid \phi(x) = \lambda x\}$ is called the *eigenspace* of λ. By Theorem 4.6.12 (4), $U(\lambda) \neq (0)$. If $v \in U(\lambda)$ and $v \neq 0$, then $\phi(v) = \lambda v$ and we say v is an *eigenvector* corresponding to λ. The reader should verify that $U(\lambda)$ is a ϕ-invariant subspace of V.

THEOREM 4.6.16. *Let k be a field, V a finite dimensional vector space over k and $\phi \in \text{Hom}_k(V,V)$. Then the eigenvalues of ϕ are precisely the roots of the minimal polynomial of ϕ.*

PROOF. Let $\lambda \in k$ and $f(x) = \min.\text{poly}_k(\phi)$. By the division algorithm, $f(x) = q(x)(x-\lambda) + f(\lambda)$. Then $f(\phi) = 0$ implies $f(\lambda) = -q(\phi)(\phi - \lambda) = -(\phi - \lambda)q(\phi)$. If λ is an eigenvalue of ϕ, then there exists a nonzero $v \in V$ such that $(\phi - \lambda)(v) = 0$. Therefore, $f(\lambda)v = 0$, which implies $f(\lambda) = 0$. Conversely, assume $f(\lambda) = 0$. Since $\deg(q) < \deg(f)$, we know $q(\phi) \neq 0$. By Theorem 4.6.12 (1), there exists $u \neq 0$ such that $v = q(\phi)u \neq 0$. Then $0 = (\phi - \lambda)q(\phi)u = (\phi - \lambda)v$. Theorem 4.6.12 (4) implies $\phi - \lambda$ is singular, hence λ is an eigenvalue of ϕ. □

THEOREM 4.6.17. *Let k be a field, V a finite dimensional vector space over k and $\phi \in \text{Hom}_k(V,V)$. Then the following are equivalent.*

1. *There is a basis B for V such that $M(\phi, B)$ is diagonal.*
2. *There is a basis of V consisting of eigenvectors of ϕ.*
3. *The minimal polynomial $\min.\text{poly}_k(\phi)$ factors into a product of linear factors in $k[x]$ and has no multiple roots.*

PROOF. (1) is equivalent to (2): This follows straight from Definitions 4.5.4 and 4.6.15.

(1) is equivalent to (3): This follows from Corollary 4.5.34. The Jordan blocks are one-by-one if and only if the exponents e_{ij} are equal to 1, if and only if the matrix is diagonal. □

EXAMPLE 4.6.18. Consider the matrix $B = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 0 & 1 & 1 \end{bmatrix}$ over the field $\mathbb{Q}$. Then $B^2 = \begin{bmatrix} 0 & 1 & 1 \\ 0 & -1 & -1 \\ -1 & 0 & 0 \end{bmatrix}$, and $B^3 = \begin{bmatrix} -1 & 0 & 0 \\ 1 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$. Using determinants we compute the characteristic polynomial of B:

$$\begin{aligned} \text{char.poly}(B) &= \det(x - B) \\ &= \begin{vmatrix} x-1 & -1 & -1 \\ 1 & x+1 & 1 \\ 0 & -1 & x-1 \end{vmatrix} \\ &= (x-1)(x+1)(x-1) + 1 + (x-1) + (x-1) \\ &= x(x^2 - x + 1). \end{aligned}$$

The roots of the characteristic polynomial are 0, $\alpha = (1-\sqrt{3}i)/2$ and $\beta = (1+\sqrt{3}i)/2$. By Theorem 4.6.16, $0, \alpha, \beta$ are also roots of the minimal polynomial of B. This proves that $\text{min. poly}(B) = x(x^2-x+1)$. The rational canonical form of B over $\mathbb{Q}$ is therefore equal to the companion matrix of $x(x^2-x+1)$, which is $C(x^3-x^2+x) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$. Let $V = \mathbb{Q}^{(3)}$ and $\psi \in \text{Hom}_{\mathbb{Q}}(V,V)$ the linear transformation corresponding to left multiplication by B. Since $\text{min. poly}(\psi)$ has degree 3, we know V is a cyclic $\mathbb{Q}[\psi]$-module. Let $u_1 = (1,0,0)^t$, $u_2 = Bu_1 = (1,-1,0)^t$, and $u_3 = Bu_2 = (0,0,-1)^t$. Then $U = \{u_1, u_2, u_3\}$ is a basis for V such that $M(\psi, U, U) = C(x^3-x^2+x)$. Set $P = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$. Then we see that $P = P^{-1}$ and $PBP = C(x^3-x^2+x)$. The Jordan canonical form of ψ exists over $F = \mathbb{Q}(\alpha)$, the splitting field of x^2-x+1. Since B has 3 distinct eigenvalues, the Jordan form of ψ is the diagonal matrix $\begin{bmatrix} 0 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \beta \end{bmatrix}$. By Theorem 4.6.17, a Jordan basis for B is a basis of eigenvectors. Using elementary row operations and gaussian elimination, the reduced row echelon form of B is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$. Therefore, $v_1 = (0,1,-1)^t$ is an eigenvector for 0. Using the identity $\alpha^2 - \alpha + 1 = 0$, we find the reduced row echelon form of $B - \alpha$ is $\begin{bmatrix} 1 & 0 & \alpha-1 \\ 0 & 1 & 1-\alpha \\ 0 & 0 & 0 \end{bmatrix}$. Therefore, $v_2 = (1-\alpha, \alpha-1, 1)^t$ is an eigenvector for α. Likewise, $v_3 = (1-\beta, \beta-1, 1)^t$ is an eigenvector for β. Then $V = \{v_1, v_2, v_3\}$ is a Jordan basis for ψ. Let P be the matrix with columns v_1, v_2, v_3. Using a symbolic calculator such as [**14**], for instance, one can show that $P^{-1}BP$ is equal to the matrix with diagonal $(0, \alpha, \beta)$.

EXAMPLE 4.6.19. Consider the matrix $A = \begin{bmatrix} 2 & 3 & 1 \\ -1 & 2 & 1 \\ 4 & -1 & -1 \end{bmatrix}$ over the field $\mathbb{Q}$. Using determinants we compute the characteristic polynomial of A:

$$\begin{aligned} \text{char. poly}(A) &= \det(x-A) \\ &= \begin{vmatrix} x-2 & -3 & -1 \\ 1 & x-2 & -1 \\ -4 & 1 & x+1 \end{vmatrix} \\ &= (x-2)^2(x+1) - 12 - 1 + (x-2) + 3(x+1) - 4(x-2) \\ &= x^2(x-3). \end{aligned}$$

The roots of the characteristic polynomial are 0, and 3. Since $A(A-3) = \begin{bmatrix} -1 & 2 & 1 \\ 3 & -6 & -3 \\ -7 & 14 & 7 \end{bmatrix}$ has rank 1, we see from Theorem 4.6.16, that the minimal polynomial of A is $\text{min. poly}(A) = x^2(x-3)$. The rational canonical form of A over $\mathbb{Q}$ is therefore equal to the companion matrix of x^3-3x^2, which is $C(x^3-3x^2) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 3 \end{bmatrix}$. Let $V = \mathbb{Q}^{(3)}$ and $\phi \in \text{Hom}_{\mathbb{Q}}(V,V)$ the

linear transformation corresponding to left multiplication by A. Since $\min.\text{poly}(\phi)$ has degree 3, we know V is a cyclic $\mathbb{Q}[\phi]$-module. Let $u_1 = (1,0,0)^t$, $u_2 = Au_1 = (2,-1,4)^t$, and $u_3 = Au_2 = (5,0,5)^t$. Then $U = \{u_1, u_2, u_3\}$ is a basis for V such that $M(\phi, U, U) = C(x^3 - 3x^2)$. Set $Q = \begin{bmatrix} 1 & 2 & 5 \\ 0 & -1 & 0 \\ 0 & 4 & 5 \end{bmatrix}$. Then we see that $AQ = QC(x^3 - 3x^2)$. The Jordan canonical form of ψ exists over $\mathbb{Q}$. By Theorem 4.3.8, the elementary divisors of ϕ are $x^2, x-3$. The Jordan canonical form for ϕ is $J(\phi) = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}$. The cyclic submodule of V corresponding to the eigenvalue 0 has dimension 2. The matrix $A - 3 = \begin{bmatrix} -1 & 3 & 1 \\ -1 & -1 & 1 \\ 4 & -1 & -4 \end{bmatrix}$ has rank 2 and $A^2(A-3) = 0$. Set $w_1 = (1,1,-4)^t$ and $w_2 = Aw_1 = (1,-3,7)^t$. Then $A^2 w_1 = 0$ and $Aw_2 = 0$. Set $w_3 = (1,0,1)^t$. Then $(A-3)w_3 = 0$, so w_3 is an eigenvector for 3. Let P be the matrix with columns w_1, w_2, w_3. The reader should verify that P is invertible and $AP = PJ(\phi)$. So w_1, w_2, w_3 is a Jordan basis for ϕ.

EXAMPLE 4.6.20. Let k be a field and $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. The characteristic polynomial of A is $(x-1)^2 - 1 = x^2 - 2x = x(x-2)$. If $\text{char}\, k \neq 2$, then A has two distinct eigenvalues, hence the Jordan form of A is diagonal: $J(A) = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}$. A Jordan basis for A is a basis of eigenvectors, $(1,-1)^t$, $(1,1)^t$. If $\text{char}\, k = 2$, then 0 is the only eigenvalue of A. The Jordan form of A is therefore $J(A) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ and a Jordan basis for A is $(1,0)^t$, $(1,1)^t$.

6.2. Exercises.

EXERCISE 4.6.21. Suppose k is a field and

$$M = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & 0 & -a_0 \\ 1 & 0 & 0 & \dots & 0 & 0 & -a_1 \\ 0 & 1 & 0 & \dots & 0 & 0 & -a_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \\ 0 & 0 & 0 & \dots & 0 & 0 & -a_{n-3} \\ 0 & 0 & 0 & \dots & 1 & 0 & -a_{n-2} \\ 0 & 0 & 0 & \dots & 0 & 1 & -a_{n-1} \end{bmatrix}$$

is a matrix in $M_n(k)$.

(1) Prove that $\min.\text{poly}_k(M) = x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$.
(2) Prove that $\text{char}.\text{poly}_k(M) = \min.\text{poly}_k(M)$.
(3) Prove that the rank of M is equal to the rank of the transpose of M.

EXERCISE 4.6.22. Let R be a commutative ring and A and B similar matrices in $M_n(R)$. Prove that $\text{char}.\text{poly}_R(A) = \text{char}.\text{poly}_R(B)$.

EXERCISE 4.6.23. Let R be a commutative ring, $A \in M_m(R)$, $B \in M_n(R)$. Define the *direct sum of A and B* by

$$A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

which is a matrix in $M_{m+n}(R)$. The direct sum $A \oplus B$ is sometimes called a *block diagonal matrix* and is denoted $\mathrm{diag}(A,B)$. Prove:

(1) $\det(A \oplus B) = \det(A)\det(B)$.
(2) $\text{char. poly}_R(A \oplus B) = \text{char. poly}_R(A)\,\text{char. poly}_R(B)$.
(3) $\mathrm{Rank}(A \oplus B) = \mathrm{Rank}(A) + \mathrm{Rank}(B)$.

EXERCISE 4.6.24. (Cramer's Rule) Let R be a commutative ring. Suppose $A \in M_n(R)$, $x, b \in R^n$ such that $Ax = b$. Prove that $x_i \det(A) = \det(B_i)$, where $B_i = (a_1, \ldots, b, \ldots, a_n)$ is the matrix obtained by replacing column i of A with the column vector b. (Hint: If $A = (a_1, \ldots, a_n)$ is written in columnar form, then $b = x_1 a_1 + \cdots + x_n a_n$. Use the multilinear and alternating properties when computing $\det(B_i)$.)

EXERCISE 4.6.25. Let $\theta : R \to S$ be a homomorphism of commutative rings.

(1) Show that θ induces a homomorphism of rings $\theta : M_n(R) \to M_n(S)$.
(2) Show that $\theta(\det(M)) = \det(\theta(M))$, for every M in $M_n(R)$.
(3) We know from Theorem 3.6.2 that θ induces a homomorphism of rings $R[x] \to S[x]$. Show that $\theta(\text{char. poly}_R(M)) = \text{char. poly}_S(\theta(M))$.

EXERCISE 4.6.26. Let $A = \begin{bmatrix} 0 & 1 & 1 \\ -4 & -4 & -1 \\ 0 & 0 & -2 \end{bmatrix}$ in the ring of 3-by-3 matrices over the field $\mathbb{Q}$.

(1) Find $\text{char. poly}(A)$, the characteristic polynomial.
(2) Find $\min. \text{poly}(A)$, the minimal polynomial.
(3) Find the invariant factors of A in $\mathbb{Q}[x]$.
(4) Find the elementary divisors of A in $\mathbb{Q}[x]$.
(5) Find the rational canonical form of A.
(6) Find the Jordan canonical form of A.
(7) Find an invertible matrix P such that $P^{-1}AP$ is equal to the Jordan canonical form of A. In other words, find a Jordan basis for the linear transformation on $\mathbb{Q}^{(3)}$ defined by A.

EXERCISE 4.6.27. Let R be a commutative ring and $A \in M_{nm}(R)$. For each i, let A_i denote column i. Assume $1 \leq i < j \leq m$ and $\alpha \in R$. If B is the matrix obtained by replacing A_j with $\alpha A_i + A_j$, show that $\det(B) = \det(A)$.

EXERCISE 4.6.28. This exercise is a generalization of Example 4.6.20. Let k be a field and $A = (a_{ij})$ the n-by-n matrix in $M_n(k)$ with $a_{ij} = 1$ for every pair (i, j).

(1) Assume the characteristic of k does not divide n. Prove the following:
   (a) $\min. \text{poly}_k(A) = x(x-n)$.
   (b) $\text{char. poly}_k(A) = \pm x^{n-1}(n-x)$.
   (c) The set

$$v_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, v_{n-1} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, v_n = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}$$

   is a Jordan basis for A.

(2) Assume the characteristic of k divides n. Prove the following:
   (a) $\min.\text{poly}_k(A) = x^2$.
   (b) $\text{char.poly}_k(A) = \pm x^n$.
   (c) The set v_1, v_2, …, v_{n-2}, $v_{n-1} = (0,0,\ldots,0,1)^t$, v_n is a Jordan basis for A, where $v_1,\ldots,v_{n-2}$ and v_n are the vectors from Part (1) (c).

EXERCISE 4.6.29. Let R be an integral domain and M a finitely generated R-module. Let $\phi \in \text{Hom}_R(M,M)$. Show that there exists a monic polynomial $p(x) \in R[x]$ such that $p(\phi) = 0$. (Hints: Exercise 4.5.26, Lemma 3.5.1, and Theorem 4.6.14 (2).)

EXERCISE 4.6.30. Let R be a commutative ring and $n \geq 1$. Define the *trace* of a matrix $\alpha = (\alpha_{ij}) \in M_n(R)$ by $\text{trace}(\alpha) = \sum_{i=1}^n \alpha_{ii}$.

(1) Prove that the trace mapping is an R-module homomorphism from $M_n(R)$ to R.
(2) Prove that $\text{trace}(\alpha\beta) = \text{trace}(\beta\alpha)$. (Hint: First show $\text{trace}(\alpha e_{ij}) = \text{trace}(e_{ij}\alpha)$ if e_{ij} is an elementary matrix and α is arbitrary.)
(3) Prove that if α and β are similar, then $\text{trace}(\alpha) = \text{trace}(\beta)$.

EXERCISE 4.6.31. Let R be a commutative ring, M a finitely generated free R-module, and X a basis for M over R. Define the trace of $\phi \in \text{Hom}_R(M,M)$ to be $\text{trace}(\phi) = \text{trace}(M(\phi,X))$. Show that this definition is independent of the choice for X. Show that the trace mapping is an R-module homomorphism from $\text{Hom}_R(M,M)$ to R.

EXERCISE 4.6.32. Let k be a field, $n \geq 1$, $f = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in k[x]$ and $M = C(f)$ the companion matrix of f. Prove the following.

(1) $\det(M) = (-1)^n a_0$.
(2) $\text{trace}(M) = -a_{n-1}$.

EXERCISE 4.6.33. Let R be a commutative ring and M a finitely generated free R-module of rank n. Let $\phi \in \text{Hom}_R(M,M)$. Show that if $\text{char.poly}_R(\phi) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$, then $\text{trace}(\phi) = -a_{n-1}$ and $\det(\phi) = (-1)^n a_0$.

EXERCISE 4.6.34. Let k be a field, V a finitely generated vector space over k, and $\phi \in \text{Hom}_k(V,V)$. Suppose $q = \min.\text{poly}_k(\phi) = x^m + a_{m-1}x^{m-1} + \cdots + a_0$ is irreducible in $k[x]$. Prove the following.

(1) $\text{char.poly}_k(\phi) = q^r$ for some integer r.
(2) $\det(\phi) = (-1)^{mr} a_0^r$.
(3) $\text{trace}(\phi) = -r a_{m-1}$.

EXERCISE 4.6.35. Let k be a field and A a matrix in $M_n(k)$ such that $\text{Rank}(A) = r < n$. Prove:

(1) $\det(A) = 0$.
(2) If B is an $r+1$-by-$r+1$ submatrix of A, then $\det(B) = 0$.
(3) A contains an r-by-r submatrix of rank r.

EXERCISE 4.6.36. Let k be a field and f an irreducible polynomial with coefficients in k. Show that if M is an n-by-n matrix over k such that $f(M) = 0$, then $\deg(f) \leq n$.

EXERCISE 4.6.37. Let R be a commutative ring and $n \geq 1$. If $A \in M_n(R)$, show that the trace of A (see Exercise 4.6.30) satisfies:

$$\sum_{i=1}^n \sum_{j=1}^n e_{ij} A e_{ji} = \text{trace}(A) I_n$$

where e_{ij} denotes the elementary matrix (Definition 4.5.2) and $I_n = e_{11} + \cdots + e_{nn}$ the identity matrix.

EXERCISE 4.6.38. Let R be a commutative ring and $A = M_n(R)$ the ring of n-by-n matrices over R. The so-called *trace pairing* $\tau : A \times A \to R$ is defined by $\tau(\alpha,\beta) = \text{trace}(\alpha\beta)$, where the trace map is defined in Exercise 4.6.30. Show that τ satisfies these properties:

(1) $\tau(\alpha,\beta) = \tau(\beta,\alpha)$.
(2) $\tau(a_1\alpha_1 + a_2\alpha_2,\beta) = a_1\tau(\alpha_1,\beta) + a_2\tau(\alpha_2,\beta)$ for $a_1, a_2 \in R$.
(3) $\tau(\alpha, b_1\beta_1 + b_2\beta_2) = b_1\tau(\alpha,\beta_1) + b_2\tau(\alpha,\beta_2)$ for $b_1, b_2 \in R$.
(4) If $\alpha \neq 0$ is fixed, then $\tau(\alpha, \ ) : A \to R$ is nonzero. That is, there exists β such that $\tau(\alpha,\beta) \neq 0$.

We say that τ is a *symmetric nondegenerate bilinear form*.

CHAPTER 5

Fields

If k is a field, there is a unique homomorphism $\eta : \mathbb{Z} \to k$ and the kernel of η is either (0), or (p) for some prime p (Example 3.2.4 (5) and Exercise 3.2.29 (2)). If η is one-to-one, then the characteristic of k is zero and k contains the quotient field of $\operatorname{im}\eta$, which is isomorphic to the field of rational numbers $\mathbb{Q}$ (Exercise 3.5.2). Otherwise, the characteristic of k is positive and the image of η is a finite field isomorphic to $\mathbb{Z}/p$, where $p = \operatorname{char} k$. The image of η is contained in every subring of k. The *prime subfield* of k is the smallest subfield P of k. Since P contains the image of η, if $\operatorname{char} k = 0$, then P is isomorphic to $\mathbb{Q}$. Otherwise, $\operatorname{char} k = p$ is positive and P is isomorphic to $\mathbb{Z}/p$. In this chapter, the study of an arbitrary field F is always in relation to its subfields. That is, F will be viewed as an extension of a subfield.

A central theme of this book is that Algebra is the study of polynomial equations. If $p(x)$ is a polynomial with coefficients over a field k, then we show in Kronecker's Theorem (see Theorem 5.2.4) that there is an extension field F of k which contains all of the roots of $p(x)$. In this chapter, groups arise as permutation groups of the roots of the polynomial $p(x)$. Since a polynomial has only a finite number of roots, in this chapter we restrict our attention to finite groups. There is a connection between the groups acting on the roots of $p(x)$ and the intermediate fields between k and F. This relationship is encapsulated in the Fundamental Theorem of Galois Theory.

1. Field Extensions

This section serves as the preparation site for the rest of the chapter. The results in Section 5.1.1 are basic and of a foundational nature. Section 5.1.1 contains an illustration of how Algebra can be applied to Geometry. Using field extensions, three questions of antiquity involving straightedge and compass constructions are answered in Theorem 5.1.17.

1.1. Algebraic Extensions and Transcendental Extensions. Let k and F be fields. If k is a subring of F, then we say F is an *extension* of k, k is a *subfield* of F, or that F/k is an *extension of fields*. An *intermediate field of* F/k is a field E such that $k \subseteq E \subseteq F$, k is a subfield of E, and E is subfield of F.

DEFINITION 5.1.1. Let F/k be an extension of fields. Then F is a k-algebra, and in particular F is a vector space over k. If $X \subseteq F$, then as in Definition 4.4.4 we denote by $k[X]$ the k-subalgebra of F generated by k and X. By $k(X)$ we denote the subfield of F generated by k and X. If $F = k(u_1, \ldots, u_n)$, then we say F is a *finitely generated field extension of* k. If $F = k(u)$, then we say F is a *simple extension* of k and u is a *primitive element*.

EXAMPLE 5.1.2. Let F be a finite field of order q. Let k be the prime subfield of F. If F has characteristic p, then k is isomorphic to $\mathbb{Z}/p$. If $\dim_k F = n$, then $q = p^n$. By Corollary 3.6.10, the group of units of F is a cyclic group of order $q-1$. Let $\zeta \in F^*$ be an element of order $q-1$. Then $F = k(\zeta)$ is a simple extension and ζ is a primitive element.

LEMMA 5.1.3. *Let F/k be an extension of fields and $X \subseteq F$.*

(1) $k[X] = \{g(u_1,\ldots,u_n) \mid n \geq 1, u_i \in X, g \in k[x_1,\ldots,x_n]\}$

(2) $k(X) = \left\{ \frac{g(u_1,\ldots,u_n)}{h(v_1,\ldots,v_n)} \mid n \geq 1, u_i, v_j \in X, g,h \in k[x_1,\ldots,x_n], h(v_1,\ldots,v_n) \neq 0 \right\}$

As k-algebras, the quotient field of $k[X]$ is isomorphic to $k(X)$.

PROOF. Is left to the reader. □

Let F/k be an extension of fields and $u \in F$. By Definition 4.4.5, u is algebraic over k if there is a nonzero polynomial $f \in k[x]$ and $f(u) = 0$. Otherwise, u is transcendental over k. If each element of F is algebraic over k, then F/k is an algebraic extension.

THEOREM 5.1.4. *(Fundamental Theorem on Algebraic Elements in a Field Extension) Let F/k be an extension of fields. Let $u \in F$ be an element that is algebraic over k. Let x be an indeterminate. The the following are true.*

(1) $k[u] = k(u)$.

(2) $k[u] \cong k[x]/(f)$ *where f is a polynomial in $k[x]$ satisfying:*

  (a) *f is monic and irreducible,*

  (b) $f(u) = 0$, *and*

  (c) *if $g \in k[x]$ and $g(u) = 0$, then $f \mid g$. The polynomial f is uniquely determined by u. We call f the* irreducible polynomial of *u and write $f = \operatorname{Irr.poly}_k(u)$. Sometimes we call f the* minimal polynomial of *u and write $f = \min.\operatorname{poly}_k(u)$.*

(3) *If $f = \operatorname{Irr.poly}_k(u)$, and $\deg f = n$, then $\{1,u,\ldots,u^{n-1}\}$ is a basis for $k[u]$ as a k-vector space.*

(4) $\dim_k k[u] = n$.

PROOF. Since u is algebraic, we know from Theorem 4.4.6 that $\deg f > 0$. If $f = gh$, then $0 = f(u) = g(u)h(u)$. Since F is a field, this implies $g(u) = 0$ or $h(u) = 0$. Theorem 4.4.6 implies that $f \mid g$ or $f \mid h$. So $\deg g = \deg f$ or $\deg h = \deg f$. This proves f is irreducible. The rest follows from Theorem 4.4.6 and Lemma 4.4.3. □

THEOREM 5.1.5. *Let F/k be an extension of fields and $u \in F$ an element that is transcendental over k. Let x be an indeterminate. Then $k(x) \cong k(u)$ by a k-algebra isomorphism that maps x to u.*

PROOF. Define $\tau : k[x] \to F$ to be the "evaluation at u" map. By Theorem 4.4.6, τ maps $k[x]$ isomorphically onto $k[u]$. By Exercise 3.5.2, τ factors through $k(x)$. Hence there is a k-algebra isomorphism $k(x) \cong k(u)$. □

THEOREM 5.1.6. *Let F/k be an extension of fields and $u \in F$. Assume L/K is another extension of fields and $v \in L$. Let $\sigma : k \to K$ be an isomorphism of fields and assume either*

(1) *u is transcendental over k and v is transcendental over K, or*

(2) *there exists an irreducible polynomial $f \in k[x]$ such that $f(u) = 0$ and $(\sigma f)(v) = 0$.*

Then there is an isomorphism $\tau : k(u) \to K(v)$ such that $\tau(u) = v$ and $\tau|_k = \sigma$.

PROOF. (1): Follows straight from Theorem 5.1.5.

(2): Because σ is an isomorphism of fields, we have an isomorphism of polynomial rings $\sigma : k[x] \to K[x]$, where $\sigma\left(\sum a_i x^i\right) = \sum \sigma(a_i) x^i$. Therefore, $\sigma(f)$ is irreducible in $K[x]$.

Then $\ker \eta\sigma = (f)$ and the diagram

$$\begin{array}{ccc} k[x] & \xrightarrow{\sigma} & K[x] \\ \downarrow & & \downarrow \eta \\ \frac{k[x]}{(f)} & \xrightarrow{\tau} & \frac{K[x]}{(\sigma f)} \end{array}$$

commutes. By the Isomorphism Theorem, τ is an isomorphism. The rest follows from Theorem 5.1.4. □

The next two corollaries play a fundamental role in Galois Theory.

COROLLARY 5.1.7. *Let F/k be an extension of fields and assume $u, v \in F$. Assume either*

(1) u and v are transcendental over k, or
(2) u and v are algebraic and satisfy the same irreducible polynomial.

Then there is a k-algebra isomorphism $\tau : k(u) \to k(v)$ such that $\tau(u) = v$.

COROLLARY 5.1.8. *Let F/k be an extension of fields. Assume $u, v \in F$ are algebraic over k and that there is a k-algebra isomorphism $\tau : k(u) \to k(v)$ such that $\tau(u) = v$. Then u and v satisfy the same irreducible polynomial.*

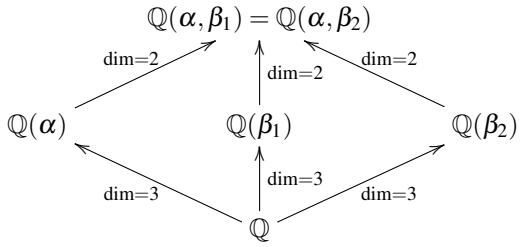
PROOF. Let $\phi : k[x] \to k[u]$ where $\phi(x) = u$. Let $\psi : k[x] \to k[v]$ where $\psi(x) = v$. The diagram of k-algebra homomorphisms

$$\begin{array}{ccc} k[x] & \xrightarrow{\phi} & k[u] \\ = \downarrow & & \downarrow \tau \\ k[x] & \xrightarrow{\psi} & k[v] \end{array}$$

commutes. Let $\ker(\phi) = (f)$, where f is the monic irreducible polynomial for u. The diagram commutes, so $f \in \ker(\psi)$. It follows that $f(v) = 0$. By Theorem 5.1.4, it follows that $\ker(\psi)$ is generated by f. □

EXAMPLE 5.1.9. In $\mathbb{Q}[x]$, let $p(x) = x^3 + 2x + 1$. By the Rational Root Theorem, $p(1) = 4$ and $p(-1) = -2$ imply $p(x)$ has no root in $\mathbb{Q}$. Therefore, p is irreducible. Since $p'(x) = 3x^2 + 2$ is positive, we see that $p(x)$ has exactly one real root, call it α. In $\mathbb{C}$ there are two nonreal roots of $p(x)$, call them β_1, β_2. Then β_1 and β_2 are complex conjugates of each other. By Corollary 5.1.7, the fields $\mathbb{Q}(\alpha), \mathbb{Q}(\beta_1), \mathbb{Q}(\beta_2)$ are pairwise isomorphic to each other. Since $\mathbb{Q}(\alpha) \subseteq \mathbb{R}$ and $\beta_i \notin \mathbb{R}$, we know that as subsets of $\mathbb{C}$, $\mathbb{Q}(\alpha)$ is not equal to $\mathbb{Q}(\beta_i)$. Therefore, over the field $\mathbb{Q}(\alpha)$, the polynomial $p(x)$ factors into $p(x) = (x - \alpha)q(x)$, where $q(x)$ is an irreducible quadratic with roots β_1, β_2. This implies $\mathbb{Q}(\alpha)(\beta_1) = \mathbb{Q}(\alpha)(\beta_2)$ has degree 2 over $\mathbb{Q}(\alpha)$. Using Galois Theory we will see later that the fields

$\mathbb{Q}(\beta_1)$ and $\mathbb{Q}(\beta_2)$ are not equal as sets.



PROPOSITION 5.1.10. *Let F/k be an extension of fields.*

1. *(Finite Dimensional is Finitely Generated and Algebraic) If F is finite dimensional over k, then F is finitely generated and algebraic over k.*
2. *(Finitely Generated and Algebraic is Finite Dimensional) If $X=\{u_1,\ldots,u_n\}\subseteq F$ and each u_i is algebraic over k, then $\dim_k k(X)<\infty$.*
3. *If $F=k(X)$ and every element of X is algebraic over k, then F is algebraic over k.*
4. *(Algebraic over Algebraic is Algebraic) Let E be an intermediate field of F/k. If F/E is algebraic and E/k is algebraic, then F/k is algebraic.*
5. *(Algebraic Closure of k in F Exists) If $E=\{u\in F \mid u \text{ is algebraic over } k\}$, then E is an intermediate field of F/k.*

PROOF. (1): Since F is finite dimensional over k, F is finitely generated (Definition 4.2.5). By Corollary 4.4.8, F is algebraic over k.

(2): By Theorem 4.4.6 (5), $\dim_k k(u_1)<\infty$. Now use induction and Proposition 4.2.8.

(3): Let $u\in k(X)$. By Lemma 5.1.3 there exist $u_1,\ldots,u_m,v_1,\ldots,v_n$ in X and polynomials f,g over k such that

$$u=\frac{f(u_1,\ldots,u_m)}{g(v_1,\ldots,v_n)}.$$

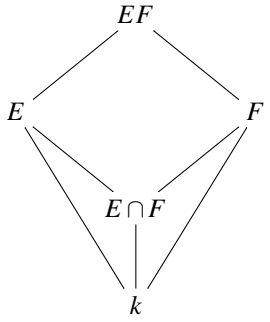
This shows $u\in k(u_1,\ldots,u_m,v_1,\ldots,v_n)$. By Parts (2) and (1) this shows u is algebraic over k.

(4): Let $u\in F$. There is a polynomial $f=\sum_{i=0}^n a_i x^i$ in $E[x]$ such that $f(u)=0$. Let $K=k(a_0,\ldots,a_n)$. Then u is algebraic over K and $\dim_K K(u)<\infty$. Since each a_i is algebraic over k, by Part (2), $\dim_k K<\infty$. By Proposition 4.2.8, $\dim_k K(u)<\infty$. By Part (1), u is algebraic over k.

(5): Let u,v be algebraic over k. By Part (3), $k(u,v)$ is an algebraic extension of k. So $k(u,v)\subseteq E$. Therefore, $u+v$, $u-v$, uv, u/v are all in E. It follows that E is a field. □

DEFINITION 5.1.11. Let K/k be an extension of fields. Let E and F be intermediate fields. That is, $k\subseteq E\subseteq K$ and $k\subseteq F\subseteq K$. The *composite* of E and F, denoted EF, is $k(E\cup F)$.

THEOREM 5.1.12. *(Fundamental Theorem on Composite Fields) Let K/k be an extension of fields. Let E and F be intermediate fields.*



Assume $\dim_k F = n$ is finite and that $\{v_1, \ldots, v_n\}$ is a basis for F as a k-vector space. The following are true.

1. *As a vector space over E, EF is spanned by $\{v_1, \ldots, v_n\}$.*
2. *$\dim_E(EF) \leq \dim_k F$.*
3. *If $\dim_k E = m$ is finite and $\{u_1, \ldots, u_m\}$ is a basis for E as a k-vector space, then $\dim_k EF \leq \dim_k E \dim_k F$ and as a vector space over k, EF is spanned by $\{u_i v_j \mid 1 \leq i \leq m, 1 \leq j \leq n\}$.*
4. *If $\dim_k E$ and $\dim_k F$ are both finite and relatively prime to each other, then $\dim_k EF = \dim_k F \dim_k E$.*
5. *If $\dim_k EF = \dim_k F \dim_k E$, then $k = E \cap F$.*

PROOF. (1): We have $F = k(v_1, \ldots, v_n)$. It follows that $EF = k(E \cup F) = k(E)(F) = E(F) = E(k(v_1, \ldots, v_n)) = E(v_1, \ldots, v_n)$. By Exercise 5.1.22, a typical element u in EF is a linear combination $u = e_1 M_1 + \cdots + e_r M_r$ where each e_i is in E and each M_i is a monomial of the form $M_i = v_1^{\varepsilon_{i,1}} \cdots v_n^{\varepsilon_{i,n}}$, where $\varepsilon_{i,j} \geq 0$ for each i, j. In the field F, each monomial M_i can be written as a k-linear combination in the form $M_i = a_{i,1} v_1 + \cdots + a_{i,n} v_n$, where $a_{i,j} \in k$ for each i, j. Therefore,

$$\begin{aligned} u &= e_1 M_1 + \cdots + e_r M_r \\ &= \sum_{i=1}^{r} e_i \left(\sum_{j=1}^{n} a_{i,j} v_j \right) \\ &= \sum_{i=1}^{r} \left(\sum_{j=1}^{n} e_i a_{i,j} v_j \right) \\ &= \sum_{j=1}^{n} \left(\sum_{i=1}^{r} e_i a_{i,j} \right) v_j \end{aligned}$$

This proves (1) since each $e_i a_{i,j}$ is in E.

(2): This follows from (1) and Corollary 4.2.6.

(3): This follows from (2) and Proposition 4.2.8.

(4): We have $\dim_k(E) = m$ and $\dim_k(F) = n$ both divide $\dim_k(EF)$. Since m and n are relatively prime, it follows that mn is the least common multiple of m and n. Thus $mn \leq \dim_k(EF)$. This and (3) proves (4).

(5): We have $\dim_k(EF) = \dim_k(F) \dim_k(E) = \dim_E(EF) \dim_k(E)$, which implies $\dim_E(EF) = \dim_k(F)$. By this and (2), $\dim_E(EF) = \dim_k(F) \leq \dim_{E \cap F}(F)$. It follows from Proposition 4.2.8 that $k = E \cap F$. □

1.2. Classical Straightedge and Compass Constructions. In this section we apply field extensions to answer three questions of antiquity on geometric constructions using straightedge and compass. The results of this section are not applied anywhere else in the book.

A real number a in $\mathbb{R}$ is *constructible* if by use of straightedge and compass we can construct a line segment of length $|a|$. We are given that 1 is constructible. Ruler and compass constructions involve:

(1) Drawing lines through two points.
(2) Intersecting two lines.
(3) Drawing a circle with a given center and radius.
(4) Intersecting a line and a circle.
(5) Intersecting two circles.

LEMMA 5.1.13. *The set of all constructible numbers is a subfield of* $\mathbb{R}$ *containing* $\mathbb{Q}$.

PROOF. Using the straightedge we can construct the x-axis. Given the unit length 1 and compass we can construct any $n \in \mathbb{Z}$. In fact, for any constructible numbers a and b, the compass can be used to construct $a \pm b$. Using the straightedge and compass we can construct the y-axis, by erecting a perpendicular to the x-axis at the number 0. The line L through the points $(0,0)$ and $(1,b)$ in $\mathbb{R}^2$ is the set of solutions to $y = bx$. The point (a,ab) is the intersection of L with the vertical line through $(a,0)$. If $b \neq 0$, the point $(a/b,b)$ is the intersection of L with the horizontal line through $(0,b)$. Therefore, ab and a/b are constructible. □

Let F be any subfield of $\mathbb{R}$. Let $F^2 = \{(x,y) \mid x,y \in F\}$ be the *plane over* F, which we view as a subset of the euclidean plan $\mathbb{R}^2$. A linear equation over F in two variables is an equation of the form $ax+by+c=0$, where a and b are in F and are not both equal to 0. A *line* in F^2 is the set of solutions $(x,y) \in F^2$ to a linear equation over F. A *circle* in F^2 is the set of solutions $(x,y) \in F^2$ to a quadratic equation of the form $x^2+y^2+ax+by+c=0$, where $a,b,c \in F$.

LEMMA 5.1.14. *The following are true.*

(1) Given $A_0 = (x_0,y_0)$ and $A_1 = (x_1,y_1)$ in F^2, if $A_0 \neq A_1$, there is a line L in F^2 passing through A_0 and A_1.
(2) Given a point $A_0 = (x_0,y_0)$ in F^2 and a positive $r \in F$, there is a circle in F^2 with center A_0 and radius r.
(3) If L_1 and L_2 are non-parallel lines in F^2, then $L_1 \cap L_2$ is a point in F^2.
(4) If L is a line and C a circle, both in F^2, and $L \cap C$ is non-empty in $\mathbb{R}^2$, then $L \cap C$ is non-empty in the plane over $F(\sqrt{\gamma})$, for some $\gamma \in F$, $\gamma \geq 0$.
(5) If C_0 and C_1 are circles in F^2, and $C_0 \cap C_1$ is non-empty in $\mathbb{R}^2$, then $C_0 \cap C_1$ is non-empty in the plane over $F(\sqrt{\gamma})$, for some $\gamma \in F$, $\gamma \geq 0$.

PROOF. (1), (2) and (3): Proofs are left to the reader.

(4): Suppose the equation for C is $x^2+y^2+ax+by+c=0$, and the equation for L is $dx+ey+f=0$, where $a,b,c,d,e,f \in F$. Without loss of generality, assume $e \neq 0$. Solve for y on the line L to get $y = -(f+dx)/e$. Substituting into C,

$$x^2 + (f+dx)^2/e^2 + ax - b(f+dx)/e + c = 0.$$

This is a quadratic equation over F of the form $Ax^2+Bx+C=0$, where $A=(e^2+d^2)/e^2>0$. In the field of complex numbers $\mathbb{C}$ the solutions are

$$x=\frac{-B\pm\sqrt{B^2-4AC}}{2A}.$$

Let $\gamma=B^2-4AC$. Then $\gamma\in F$. If $\gamma=0$, then $L\cap C$ consists of one point in F^2. If $\gamma<0$, then in $\mathbb{R}^2$, $L\cap C=\emptyset$. If $\gamma>0$, then there are two points in $L\cap C$, and both belong to the plane over $F(\sqrt{\gamma})$.

(5): Suppose the equation for C_0 is $x^2+y^2+a_0x+b_0y+c_0=0$, and the equation for C_1 is $x^2+y^2+a_1x+b_1y+c_1=0$. If $C_0=C_1$, then take γ to be 1. Otherwise subtract to get $(a_0-a_1)x+(b_0-b_1)y+(c_0-c_1)=0$. If $a_0=a_1$ and $b_0=b_1$, then $C_0\cap C_1=\emptyset$. Otherwise the linear equation $(a_0-a_1)x+(b_0-b_1)y+(c_0-c_1)=0$ defines a line, which we call L. Then $C_0\cap L=C_1\cap L=C_0\cap C_1$, and we reduce to part (4). □

PROPOSITION 5.1.15. *If $u\in\mathbb{R}$ is constructible, then for some $r\geq 0$, $\dim_{\mathbb{Q}}(\mathbb{Q}(u))$ is equal to 2^r.*

PROOF. To construct u, a finite sequence of straightedge and compass constructions are performed. By Lemma 5.1.14, u belongs to a field extension of $\mathbb{Q}$ obtained by a finite number of quadratic extensions, each of which is inside $\mathbb{R}$. There exist positive real numbers $\gamma_1,\ldots,\gamma_n$ such that u belongs to $\mathbb{Q}(\gamma_1)\cdots(\gamma_n)$, a subfield of $\mathbb{R}$. Moreover, $\gamma_1^2\in\mathbb{Q}$ and for $1<i\leq n$, $\gamma_i^2\in\mathbb{Q}(\gamma_1,\ldots,\gamma_{i-1})$. By Proposition 4.2.8, degrees of consecutive extensions multiply. The degree of each consecutive extension is either 1 or 2. This means $\dim_{\mathbb{Q}}(\mathbb{Q}(\gamma_1,\ldots,\gamma_n))$ is 2^s for some $s\geq 0$. Since $\dim_{\mathbb{Q}}(\mathbb{Q}(u))$ divides 2^s, we are done. □

COROLLARY 5.1.16. *Suppose $u\in\mathbb{R}$ is algebraic over $\mathbb{Q}$ and the degree of* $\mathrm{Irr.poly}_{\mathbb{Q}}(u)$ *has degree d. If d is not of the form 2^r, then u is not constructible.*

THEOREM 5.1.17. *It is impossible by straightedge and compass alone to*

(1) trisect the angle 60° (that is, $\cos 20^\circ$ is not constructible),
(2) double the cube (that is, $\sqrt[3]{2}$ is not constructible), or
(3) square the circle (that is, $\sqrt{\pi}$ is not constructible).

PROOF. (1): Take θ to be 60°. Then $\cos\theta=\frac{1}{2}$. By trigonometry, $\cos\theta=4\cos^3\left(\frac{\theta}{3}\right)-3\cos\left(\frac{\theta}{3}\right)$. Let $u=2\cos 20^\circ$. Then u satisfies $u^3-3u-1=0$. The irreducible polynomial for u over $\mathbb{Q}$ is x^3-3x-1, which has degree 3. Then u is not constructible, $\cos 20^\circ$ is not constructible, and it is impossible to trisect 60°.

(2): The irreducible polynomial for $\sqrt[3]{2}$ over $\mathbb{Q}$ is x^3-2, which has degree 3.

(3): We have not proved it here, but π is transcendental. Hence $\sqrt{\pi}$ is not constructible. □

1.3. Exercises.

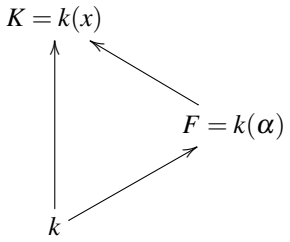
EXERCISE 5.1.18. Let p be an odd prime and $k=\mathbb{Z}/p$ the field of order p. Show that there are $(p-1)/2$ elements $\alpha\in U_p$ such that $\phi_\alpha=x^2-\alpha$ is irreducible. Show that in this case $k[x]/(\phi_\alpha)$ is a field of order p^2.

EXERCISE 5.1.19. Let $k=\mathbb{Z}/3$ be the field of order 3. Show that $f=x^2+1$ is irreducible over k. Let $F=k[x]/(f)$. Let $u\in F$ be the coset represented by x. By Corollary 3.6.10, the group F^* is cyclic. A generator for F^* is called a primitive element. Show that $u+1,u-1,-u+1,-u-1$ are the four primitive elements in F^*.

EXERCISE 5.1.20. Let $p(x) = x^3 - 3x - 1 \in \mathbb{Q}[x]$. Show that p is irreducible and let $F = \mathbb{Q}[x]/(p)$ be the quotient. Let u denote the element of F corresponding to the coset containing x.

(1) Exhibit a basis for F as a $\mathbb{Q}$-vector space.
(2) Write the following in terms of the basis given in (1): u^{-1}, $u^4 + 2u^3 + 3$, u^{-2}.

EXERCISE 5.1.21. Let k be a field, x an indeterminate, and $K = k(x)$ the field of rational functions. Let α denote the rational function $x^4/(4x^3 - 1)$ in K. Then $F = k(\alpha)$ is a field extension of k and K is a field extension of F. There is a lattice of subfields



where an arrow denotes set containment. Show that K is algebraic over F. Determine the minimal polynomial of x over F and the dimension $\dim_F(K)$. (Hint: Apply Theorem 3.7.9 to show that $y^4 - \alpha(4y^3 - 1)$ is an irreducible polynomial in $K[y]$.)

EXERCISE 5.1.22. Let K/k be an extension of fields and $u_1, \ldots, u_n$ elements of K, where $n \geq 1$. As in Definition 5.1.1, $k[u_1, \ldots, u_n]$ is the k-subalgebra of K generated by k and $u_1, \ldots, u_n$. Show that a typical element in $k[u_1, \ldots, u_n]$ can be written as a sum of the form $k_1 M_1 + \cdots + k_r M_r$ where $a_i \in k$ for each i and each M_i is a product of the form $M_i = u_1^{\varepsilon_{i,1}} \cdots u_n^{\varepsilon_{i,n}}$ where $\varepsilon_{i,j} \geq 0$ for each i, j.

EXERCISE 5.1.23. Let F/k be a finite dimensional extension of fields. If E is an intermediate field of F/k, show that F/E is finite dimensional, E/k is finite dimensional, and $\dim_k(F) = \dim_k(E)\dim_E(F)$.

2. Algebraic Field Extensions

There are two main results in this section. Let k be a field and f a polynomial over k. The main result of Section 5.2.1 is the proof that there is a unique extension F/k generated by adjoining the roots of f to k (Corollary 5.2.8). The main result of Section 5.2.2 is the Primitive Element Theorem (Theorem 5.2.14) which contains sufficient conditions for an algebraic extension of fields to be a simple extension.

2.1. Existence and Uniqueness of a Splitting Field. Let k be a field and p a polynomial in $k[x]$ of positive degree. If F/k is an extension of fields, then we say that p *splits* in F if each irreducible factor of p in $F[x]$ is linear. Equivalently, p factors in $F[x]$ into a product of linear polynomials.

LEMMA 5.2.1. *Let F be a field. The following are equivalent.*

(1) Every nonconstant polynomial $p \in F[x]$ has a root in F.
(2) Every nonconstant polynomial $p \in F[x]$ splits in F.
(3) Every irreducible polynomial $p \in F[x]$ has degree 1.
(4) If K/F is an algebraic extension of fields, then $F = K$.
(5) F contains a subfield k such that F/k is algebraic and every polynomial in $k[x]$ splits in F.

PROOF. (1), (2), and (3) are clearly equivalent.

(2) implies (5): Is trivial.

To show (3) and (4) are equivalent, use Theorem 5.1.4.

(5) implies (4): If K/F is algebraic, then by Proposition 5.1.10 (4), K/k is algebraic. If $u \in K$, then the irreducible polynomial of u over k splits in F. Therefore $u \in F$. □

DEFINITION 5.2.2. If F is a field that satisfies any of the equivalent statements of Lemma 5.2.1, then we say F is *algebraically closed*. If F/k is an extension of fields, we say F is an *algebraic closure* of k in case F is algebraic over k, and F is algebraically closed.

DEFINITION 5.2.3. Let F/k be an extension of fields and p a nonconstant polynomial in $k[x]$. We say that F is a *splitting field* of p if

(1) p splits in F, and
(2) $F = k(u_1, \ldots, u_n)$ where $p(u_i) = 0$ for each i.

THEOREM 5.2.4. *(Kronecker's Theorem) Let k be a field and f a polynomial of positive degree in $k[x]$. There exists an extension field F of k and an element $u \in F$ satisfying*

(1) u is a root of f,
(2) $\dim_k(k[u]) \leq \deg(f)$, and
(3) if f is irreducible, then $\dim_k(k[u]) = \deg(f)$ and $k[u]$ is unique up to a k-algebra isomorphism.

PROOF. Let p be an irreducible factor of f. Write $f = pq$. Let $F = k[x]/(p)$ and take u to be the coset represented by x in F. Then $p(u) = p([x]) = [p(x)] = [0]$. Then $f(u) = p(u)q(u) = 0$. The rest follows from Theorems 5.1.4 and 5.1.6. □

EXAMPLE 5.2.5. Let p be a prime and k a field of characteristic p. Let $\alpha \in k$ and $f = x^p - \alpha$. In this example we show that f is either irreducible, or splits. The Frobenius homomorphism $\theta : k \to k$ is defined by $a \mapsto a^p$ (Exercise 3.2.31). If $\alpha = a^p$ for some $a \in k$, then $f = x^p - a^p = (x-a)^p$ by (Exercise 3.2.30). This shows that f splits over k if f has a root in k. Now assume that α is not in the image of the Frobenius map. Thus f does not have a root in k. For sake of contradiction assume f is reducible over k. Let $f = gg_1$ where g is irreducible and $\deg g = m$ where $1 \leq m < p$. Let $F = k[x]/(g)$. By Theorem 5.2.4, F is an extension field of k containing a root u of g. Every root of g is a root of f. By the first part, $f = (x-u)^p$ in $F[x]$. By Corollary 3.6.5, $F[x]$ is a UFD. This implies $g = (x-u)^m$ in $F[x]$. But $g \in k[x]$. By the Binomial Theorem, $g = x^m - mux^{m-1} + \cdots + (-u)^m$, which implies $mu \in k$. But $\gcd(m, p) = 1$ implies $u \in k$. This contradicts our original assumption that f does not have a root in k. We have shown that $f = x^p - \alpha$ is either irreducible, or splits.

COROLLARY 5.2.6. *If k is a field and f a polynomial in $k[x]$ of positive degree n, then there exists a splitting field F/k for f such that $\dim_k(F) \leq n!$.*

PROOF. Factor $f = p_1 \ldots p_m$ in $k[x]$ where each p_i is irreducible. If $\deg p_i = 1$ for each i, then take $F = k$ and stop. Otherwise, assume $\deg p_1 > 1$. By Kronecker's Theorem (Theorem 5.2.4), there is an extension field F_1/k such that $F_1 = k(\alpha)$ and $p_1(\alpha) = 0$. Note that $f(\alpha) = 0$ and $\dim_k(F_1) = \deg p_1 \leq n$. Factor $f = (x-\alpha)g$ in $F_1[x]$. By induction on n, there exists a splitting field F/F_1 for g and $\dim_{F_1}(F) \leq (n-1)!$. So f splits in F and there exist roots $u_1, \ldots, u_m$ of f such that $F = F_1(u_1, \ldots, u_m) = k(\alpha, u_1, \ldots, u_m)$. Lastly, $\dim_k(F) = \dim_k(F_1)\dim_{F_1}(F) \leq n!$, by Proposition 4.2.8. □

LEMMA 5.2.7. *Let k be a field, f a polynomial in $k[x]$ of positive degree n, and F a splitting field for f over k. Let $\sigma : k \to K$ be an isomorphism of fields, $\sigma(f)$ the image of f in $K[x]$. Let L/K be an extension field such that $\sigma(f)$ splits in L. Then σ extends to a homomorphism of k-algebras $\bar{\sigma} : F \to L$ making a commutative*

$$\begin{array}{ccc} F & \xrightarrow{\bar{\sigma}} & L \\ \uparrow & & \uparrow \\ k & \xrightarrow{\sigma} & K \end{array}$$

diagram. Every root of f in F is mapped by $\bar{\sigma}$ to a root of $\sigma(f)$ in L. If L is a splitting field for $\sigma(f)$, then $\bar{\sigma}$ is an isomorphism.

PROOF. If $F = k$, then take $\bar{\sigma} = \sigma$ and stop. Otherwise, $\dim_k(F) > 1$ and there is an irreducible factor g of f such that $\deg g > 1$. Let α be a root of g in F and β a root of $\sigma(g)$ in L. By Theorem 5.1.6 there is a k-algebra isomorphism $\tau : k(\alpha) \to K(\beta)$ such that $\tau(\alpha) = \beta$ and the bottom square of the diagram

$$\begin{array}{ccc} F & \overset{\exists\bar{\sigma}}{\dashrightarrow} & L \\ \uparrow & & \uparrow \\ k(\alpha) & \xrightarrow[\cong]{\tau} & K(\beta) \\ \uparrow & & \uparrow \\ k & \xrightarrow[\cong]{\sigma} & K \end{array}$$

commutes. Also, F is a splitting field for f over $k(\alpha)$, and $\dim_{k(\alpha)}(F) < \dim_k(F)$. By induction on $\dim_k(F)$, τ can be extended to a k-algebra homomorphism $\bar{\sigma} : F \to L$ such that the entire diagram above commutes. A root of f is mapped under $\bar{\sigma}$ to a root of $\sigma(f)$. Since f splits in F, $\sigma(f)$ splits in $\bar{\sigma}(F)$. The polynomial $\sigma(f)$ has at most $\deg(f)$ roots in L by Corollary 3.6.8, and they all belong to $\bar{\sigma}(F)$. If $\lambda \in L$ is a root of $\sigma(f)$, then $\lambda \in \bar{\sigma}(F)$. If L/K is generated by roots of $\sigma(f)$, then $L \subseteq \bar{\sigma}(F)$ and $\bar{\sigma}$ is an isomorphism. □

COROLLARY 5.2.8. *Let k be a field and $f \in k[x]$. A splitting field for f exists and is unique up to k-algebra isomorphism.*

PROOF. This follows straight from Corollary 5.2.6 and Lemma 5.2.7. □

EXAMPLE 5.2.9. Let $n \geq 2$. In $\mathbb{C}$, let $\zeta = e^{2\pi i/n}$. Then ζ is a primitive nth root of unity. That is, $\{\zeta^k \mid 0 \leq k \leq n-1\}$ are the n distinct roots of $x^n - 1$ in $\mathbb{C}$. Therefore, in $\mathbb{C}[x]$

$$x^n - 1 = (x-1)(x-\zeta)(x-\zeta^2)\cdots(x-\zeta^{n-1})$$

is the unique factorization of $x^n - 1$. For each k, $\zeta^k \in \mathbb{Q}(\zeta)$. This shows that $\mathbb{Q}(\zeta)$ is a splitting field for $x^n - 1$ over $\mathbb{Q}$. Consider the polynomial

$$\Phi_n(x) = 1 + x + \cdots + x^{n-1} = \frac{x^n - 1}{x - 1}$$

of degree $n-1$. The distinct roots of Φ_n in $\mathbb{C}$ are $\zeta, \zeta^2, \ldots, \zeta^{n-1}$. By the same reasoning as above, $\mathbb{Q}(\zeta)$ is a splitting field for Φ_n over $\mathbb{Q}$. If p is a prime, then by Example 3.7.8, Φ_p is irreducible over $\mathbb{Q}$. By Theorem 5.1.4, $\Phi_p = \text{Irr.poly}_{\mathbb{Q}}(\zeta)$, $\mathbb{Q}(\zeta) = \mathbb{Q}[x]/(\Phi_p)$, and

$\{1, \zeta, \zeta^2, \ldots, \zeta^{p-2}\}$ is a basis for $\mathbb{Q}(\zeta)$ as a $\mathbb{Q}$-vector space. The polynomial $\Phi_p(x)$ is called the *pth cyclotomic polynomial*.

2.2. The Primitive Element Theorem. Let k be a field, $f \in k[x]$, and F/k a splitting field for f. By Corollary 5.2.8, F exists and is unique up to a k-algebra isomorphism. We say f is *separable* in case for every irreducible factor p of f, every root of p in F is a simple root. If K/k is a extension of fields, then we say K/k is a *separable extension* if every $u \in K$ is the root of a separable polynomial in $k[x]$. If $u \in K$ is the root of a separable polynomial in $k[x]$, then we say u is *separable*. A separable extension is an algebraic extension. If $\operatorname{char} k = 0$, then by Theorem 3.6.15 (1), every polynomial $f \in k[x]$ is separable. The purpose of this section is to prove Theorem 5.2.14 which shows that a finite separable extension of fields is a simple extension.

EXAMPLE 5.2.10. Let k be a field of prime characteristic p. The Frobenius homomorphism $\theta : k \to k$ is defined by $a \mapsto a^p$ (Exercise 3.2.31). The image of θ is denoted k^p. Assume θ is not onto and let $\alpha \in k - k^p$. As shown in Example 5.2.5, the polynomial $f = x^p - \alpha$ is irreducible in $k[x]$ but is not separable.

LEMMA 5.2.11. *Let k be a field and f an irreducible polynomial in $k[x]$. The formal derivative of f is denoted f' (see Definition 3.6.13).*

(1) The following are equivalent:
  (a) f is separable.
  (b) $\gcd(f, f') = 1$.
  (c) $f' \neq 0$.

(2) If f is not separable, then $\operatorname{char} k = p$ is a prime number and there exists a polynomial $g(x) \in k[x]$ such that $f(x) = g(x^p)$.

PROOF. This follows from Theorem 3.6.15. □

THEOREM 5.2.12. *Let F be a finite field with $\operatorname{char} F = p$. Let k be the prime subfield of F and $n = \dim_k(F)$.*

(1) The group of units of F is a cyclic group.
(2) $F = k(u)$ is a simple extension, for some $u \in F$.
(3) The order of F is p^n.
(4) F/k is a separable extension.
(5) F is the splitting field for the separable polynomial $x^{p^n} - x$ over k.
(6) Any two finite fields of order p^n are isomorphic as fields.

PROOF. As a k-vector space, F is isomorphic to k^n, which has cardinality $|k|^n$, by Exercise 1.1.12. By Corollary 3.6.10, the group of units of F is a finite cyclic group of order $p^n - 1$. If u is a generator for F^*, then $F = k(u)$. The polynomial $x^{p^n} - x = x(x^{p^n-1} - 1)$ has p^n distinct roots in F. Therefore F is the splitting field for the separable polynomial $x^{p^n} - x$ over k and every element of F is separable over k. By Corollary 5.2.8, F is unique up to k-algebra isomorphism. □

LEMMA 5.2.13. *Let F/k be an extension of fields. Let α and β be elements of F that are algebraic over k. If β is separable over k, then there exists $\gamma \in F$ such that $k(\alpha, \beta) = k(\gamma)$.*

PROOF. First we prove the lemma for some special cases. Let $K = k(\alpha, \beta)$. If $\alpha \in k$, then $K = k(\beta)$, so set $\gamma = \beta$. If $\beta \in k$, then $K = k(\alpha)$, so set $\gamma = \alpha$. If k is a finite field, then K is a finite field by Proposition 5.1.10 (2). In this case $K = k(\gamma)$ is a simple extension, by

Theorem 5.2.12 (2). Assume from now on that $\alpha \notin k$, $\beta \notin k$, and k is infinite. The proof of the general case is split into a sequence of three steps.

Step 1 is to define a candidate for γ. Let $f = \min.\text{poly}_k(\alpha)$ and $g = \min.\text{poly}_k(\beta)$. Let F_1 be a splitting field for fg over F. Let $\alpha = \alpha_1, \alpha_2, \ldots, \alpha_m$ be the distinct roots of f in F_1. Let $\beta = \beta_1, \beta_2, \ldots, \beta_n$ be the distinct roots of g in F_1. By our hypotheses, $m \geq 2$ and $n \geq 2$. Consider

$$S = \left\{ \frac{\alpha_1 - \alpha_i}{\beta_j - \beta_1} \mid i = 2, \ldots, m \text{ and } j = 2, \ldots, n \right\}$$

which is a finite subset of F. Since k is infinite, there exists $c \in k^*$ such that $c \notin S$. Set $\gamma = \alpha + c\beta$. So $c \in k(\alpha, \beta)$. To finish, it is enough to show $\alpha \in k(\gamma)$ and $\beta \in k(\gamma)$.

Step 2 is to show that $\gamma = \alpha_i + c\beta_j$ if and only if $i = j = 1$. If $1 \leq i \leq m$ and $1 \leq j \leq n$ and $\gamma = \alpha_i + c\beta_j$, then $\alpha_i + c\beta_j = \alpha + c\beta$. So $c(\beta_j - \beta) = \alpha - \alpha_i$. If $j = 1$, then $i = 1$. If $j \neq 1$, then $c = (\alpha - \alpha_i)/(\beta_j - \beta)$. This contradicts the choice of c. This completes Step 2.

Step 3 is to show that $k(\alpha, \beta) \subseteq k(\gamma)$. Define $h(x) \in k(\gamma)[x]$ by $h(x) = f(\gamma - cx)$. Then $h(\beta) = f(\gamma - c\beta) = f(\alpha) = 0$. If $j > 1$, then $\gamma - c\beta_j \neq \alpha_i$ for any i. Thus $h(\beta_j) = f(\gamma - c\beta_j) \neq 0$. Thus, $\beta_2, \ldots, \beta_n$ are not roots of $h(x)$. Let $g_1 = \min.\text{poly}_{k(\gamma)}(\beta)$. Since $h(\beta) = 0$, by Theorem 5.1.4 we know $g_1 \mid h$. Likewise, $g(\beta) = 0$ implies $g_1 \mid g$. Every root of g_1 is a root of h and g. We proved that the only root g and h have in common is β. At this point in the proof we use the fact that g is separable. It follows that $\gcd(g, h) = x - \beta$. Hence g_1 is linear with one root, β, which implies $\beta \in k(\gamma)$. Moreover, $\alpha = \gamma - c\beta \in k(\beta, \gamma) = k(\gamma)$. □

THEOREM 5.2.14. *(The Primitive Element Theorem) Let F/k be a finite dimensional separable extension of fields. Then there is a separable element $u \in F$ such that $F = k(u)$.*

PROOF. Let $\dim_k(F) = n$. Let $\alpha_1, \ldots, \alpha_n$ be a basis for F as a k-vector space. For $i = 1, \ldots, n$, let $F_i = k(\alpha_1, \ldots, \alpha_i)$. Then So $F_2 = k(\alpha_1, \alpha_2)$. Lemma 5.2.13 implies there exists $\gamma_2 \in F$ such that $F_2 = k(\gamma_2)$. By the same argument, $F_3 = F_2(\alpha_3) = k(\gamma_2, \alpha_3)$ and there exists $\gamma_3 \in F$ such that $F_3 = k(\gamma_3)$. Iterate this process $n - 1$ times. Hence $F = F_n = k(\gamma_n)$ for some γ_n. □

2.3. Exercises.

EXERCISE 5.2.15. Show that two finite fields E and F are isomorphic if and only if the order of E is equal to the order of F.

EXERCISE 5.2.16. Let $\alpha = \sqrt[3]{2}$ be the cube root of 2 in $\mathbb{R}$ and $\zeta = e^{2\pi i/3}$ a primitive cube root of 1 in $\mathbb{C}$.

(1) Show that the splitting field for $x^3 - 2$ over $\mathbb{Q}$ is $\mathbb{Q}(\zeta, \alpha)$ and that $\dim_{\mathbb{Q}} \mathbb{Q}(\zeta, \alpha) = 6$.
(2) Show that $\mathbb{Q}(\zeta, \alpha)$ is equal to the composite field EF where E and F are any two fields from this list: $\mathbb{Q}(\zeta)$, $\mathbb{Q}(\alpha)$, $\mathbb{Q}(\zeta\alpha)$, $\mathbb{Q}(\zeta^2\alpha)$.
(3) Show that $\text{Irr.poly}_{\mathbb{Q}(\zeta)}(\alpha)$ has degree 3. Show that $\text{Irr.poly}_{\mathbb{Q}(\zeta)}(\zeta\alpha)$ has degree 3. Show that $\text{Irr.poly}_{\mathbb{Q}(\zeta)}(\zeta^2\alpha)$ has degree 3.
(4) Show that $\text{Irr.poly}_{\mathbb{Q}(\zeta\alpha)}(\alpha)$ has degree 2. Show that $\text{Irr.poly}_{\mathbb{Q}(\zeta^2\alpha)}(\alpha)$ has degree 2.

EXERCISE 5.2.17. Let F/k be an extension of fields and assume $\dim_k F = p$ is prime. Let u be any element of F that is not in k. Prove that $F = k(u)$.

EXERCISE 5.2.18. Let F/k be an extension of fields and assume $\dim_k F = 2$. Let u be an element of F that is not in k and $f = \text{Irr.poly}_k u$. Show that F is a splitting field for f over k.

EXERCISE 5.2.19. Let K/k be an extension of fields. Let F_1, F_2 be two intermediate fields where $k \subseteq F_i \subseteq K$ and $\dim_k F_i = 2$ for each i. Suppose there exists a k-algebra isomorphism $\sigma : F_1 \to F_2$. Show that F_1 and F_2 are equal as sets.

EXERCISE 5.2.20. Let $k = \mathbb{F}_2$ be the field of order 2. In $k[x]$, let $f = x^2$, $g = x^2 + 1$, and $h = x^2 + x + 1$. Show that the following four rings are distinct in the sense that no two are isomorphic to each other: $\mathbb{Z}/(4)$, $k[x]/(f)$, $k[x]/(g)$, $k[x]/(h)$. For a continuation of this exercise, see Exercise 5.6.10.

EXERCISE 5.2.21. Let k be a field and A a finite dimensional k-algebra. Prove that if $\dim_k(A) = 2$, then A is commutative.

EXERCISE 5.2.22. True or False. Justify your answers.

(1) $\mathbb{Q}(\sqrt{2}) \cong \mathbb{Q}(\sqrt{3})$
(2) $\mathbb{R}(\sqrt{-2}) \cong \mathbb{R}(\sqrt{-3})$

EXERCISE 5.2.23. Let k be a field and $K = k(x)$ the field of rational functions over k in the variable x. Let $\sigma : K \to K$ be the function which maps a typical rational function $f(x) \in K$ to the rational function $f(x^{-1})$. Show that σ is an automorphism of the field K. (Hint: Corollary 5.1.7.)

EXERCISE 5.2.24. This exercise is a continuation of Exercise 4.5.54. Let k be a field and A a matrix in $M_n(k)$. Prove that A is similar to the transpose of A.

3. Galois Theory

Let k be a field, $f \in k[x]$ a separable polynomial, and F a splitting field for f over k. The roots of f are the solutions to the algebraic equation $f(x) = 0$. The field extension F/k is generated by the roots of f. As in Definition 4.4.1, by $\text{Aut}_k(F)$ we denote the group of all k-algebra automorphisms of F. In Theorem 5.3.15 we show that F/k is a so-called Galois extension. For a Galois extension, the group $\text{Aut}_k(F)$ acts not only on F, but on the set of roots of $f(x)$. Moreover the action of the group $\text{Aut}_k(F)$ on F is entirely determined by its action on the roots of $f(x)$. In the Fundamental Theorem of Galois Theory (Theorem 5.3.18), we show that there is a one-to-one correspondence between the intermediate fields of F/k and the subgroups of $\text{Aut}_k(F)$. By this theorem, the study of the roots of the polynomial equation $f(x) = 0$ is reduced to the study of the action of a finite group acting on the set of roots. It was Galois himself who emphasized the importance of studying the set of roots of a polynomial under the action by a finite group of permutations (see [**6**]).

3.1. A Group Acting on a Field. In this section we will be using some results as well as some terminology from Group Theory. For instance, if a group of permutations G acts on a set X, there is the well defined notion of the subset of X fixed by G. Also, for any subset S of X there is the subgroup of G fixing S. The reader is referred to Section 2.4.1, especially Definition 2.4.9. While the underlying theory applies, the notation and terminology in the present context are slightly different than that of Chapter 2. Proposition 5.3.1 extends to the context of field extensions these important notions from Group Theory.

PROPOSITION 5.3.1. *Let F/k be an extension of fields and $G = \text{Aut}_k(F)$.*

(1) If H is a subset of G, then

$$F^H = \{v \in F \mid \sigma(v) = v \text{ for all } \sigma \in H\}$$

is an intermediate field of F/k which is called the fixed field of H.

(2) If E is an intermediate field of F/k, then

$$G_E = \{\sigma \in G \mid \sigma(v) = v \text{ for all } v \in E\}$$

is a subgroup of G which is called the subgroup of G fixing E. Note that $G_E = \mathrm{Aut}_E(F)$.

PROOF. The proof is left to the reader. □

PROPOSITION 5.3.2. *Let F/k be an extension of fields.*

(1) Let $f \in k[x]$, $\sigma \in \mathrm{Aut}_k(F)$, and $u \in F$. If $f(u) = 0$, then $f(\sigma(u)) = 0$.

(2) Assume $u \in F$ is algebraic over k and $E = k[u]$. If $\sigma \in \mathrm{Aut}_k(E)$, then σ is completely determined by $\sigma(u)$.

PROOF. (1): If $f = \sum_{i=0}^{n} a_i x^i$, then

$$f(\sigma(u)) = \sum_{i=0}^{n} a_i(\sigma(u))^i = \sum_{i=0}^{n} \sigma(a_i u^i) = \sigma(\sum_{i=0}^{n} a_i u^i) = \sigma(f(u)) = \sigma(0) = 0.$$

(2): By Theorem 5.1.4, there is a k-basis for E of the form $1, u, u^2, \ldots, u^{n-1}$ where $n = \dim_k(E)$. □

EXAMPLE 5.3.3. Let $\mathbb{F}_2 = \{0,1\}$ be the field of order 2, which is isomorphic to the ring $\mathbb{Z}/2$. Let $p(x) = x^2 + x + 1 \in \mathbb{F}_2[x]$. Since $p(0) = p(1) = 1$, $p(x)$ has no root in $\mathbb{F}_2$ and is irreducible in $\mathbb{F}_2[x]$. Let F be the splitting field of $p(x)$. Then F has order 4. Let α be a root of $p(x)$ in F. Then $\alpha^2 = \alpha + 1$ and by Theorem 5.1.4, $F = \{0, 1, \alpha, \alpha + 1\}$. By Theorem 5.2.12, F is unique up to isomorphism. Let $\phi \in \mathrm{Aut}(F)$. Then $\phi(0) = 0$, $\phi(1) = 1$ and $\phi(\alpha)$ is equal to α or $\alpha + 1$. If $\phi(\alpha) = \alpha$, then ϕ is equal to $1 \in \mathrm{Aut}(F)$, the identity function. By Proposition 5.3.2, ϕ is determined by the value of $\phi(\alpha)$. Therefore, $\mathrm{Aut}(F)$ has order at most 2. We prove that there is an automorphism of order two in $\mathrm{Aut}(F)$. By Exercise 3.2.31, the Frobenius homomorphism $\sigma : F \to F$ defined by $\sigma(a) = a^2$ is a homomorphism. Since F is a finite field, σ is necessarily one-to-one and onto (Exercises 3.2.28 and 1.1.11). Since $\sigma(\alpha) = \alpha^2 = \alpha + 1$, we have shown that $\mathrm{Aut}(F)$ has order two.

EXAMPLE 5.3.4. The polynomial $p(x) = x^2 + 1$ is irreducible in $\mathbb{Q}[x]$. The roots of $p(x)$ in $\mathbb{C}$ are $i, -i$. Let $F = \mathbb{Q}(i) = \mathbb{Q}(i)$ be the splitting field for $p(x)$ over $\mathbb{Q}$. By Theorem 5.1.4, a basis for F over $\mathbb{Q}$ is $1, i$. By Corollary 5.1.7, there exists an automorphism $\chi : \mathbb{Q}(i) \to \mathbb{Q}(i)$ such that $\chi(i) = -i$. The automorphism χ is usually called *complex conjugation* (see Section 1.4). By Proposition 5.3.2 (1), if $\phi \in \mathrm{Aut}_{\mathbb{Q}}(F)$, then $\phi(i)$ is equal to either i or $-i$. By Proposition 5.3.2 (2), this implies $\mathrm{Aut}_{\mathbb{Q}}(F)$ has order at most two. This proves $\mathrm{Aut}_{\mathbb{Q}}(F) = \langle\sigma\rangle$ is a cyclic group of order two.

In the next theorem we show that automorphisms of a field are linearly independent over F. In Section 5.3 it will not be necessary to view the automorphisms as elements in an F-vector space, but Theorem 5.3.5 will be applied in the proof of Lemma 5.5.1 where we show that if F/k is an extension of fields, then the endomorphism ring $\mathrm{Hom}_k(F,F)$ is an F-vector space which contains the automorphism group $\mathrm{Aut}_k(F)$ as a subset.

Theorems 5.3.5, 5.3.6, and 5.3.8 play key roles in Galois Theory. The proof we give below of the Fundamental Theorem (Theorem 5.3.18) relies heavily on these three theorems. The statements of the theorems concern a field and its group of automorphisms. Because the topic of this section is Galois Theory, this is to be expected. But, upon a close look at the proofs of these results, the common feature that stands out is that they are strictly of a linear algebra nature. From this perspective, we see that Galois Theory is inherently a part of Linear Algebra.

THEOREM 5.3.5. *Let F be a field and $\sigma_1, \ldots, \sigma_n$ a finite set of distinct automorphisms of F. If $u_1, \ldots, u_n$ are elements of F and*

$$u_1\sigma_1(x) + \cdots + u_n\sigma_n(x) = 0 \tag{3.1}$$

for all $x \in F$, then each u_i is equal to zero.

PROOF. For sake of contradiction assume a nontrivial dependence relation of the type (3.1) exists. Pick one such relation involving a minimal number of the automorphisms. If necessary, relabel the automorphisms and assume

$$u_1\sigma_1 + \cdots + u_r\sigma_r = 0 \tag{3.2}$$

where $u_1, \ldots, u_r$ are all nonzero and r is minimal. Since $\sigma_i(1) = 1$ for each i, in (3.2) we have $r \geq 2$. For some $y \in F$ we have $\sigma_1(y) \neq \sigma_r(y)$. Evaluating (3.2) at yx, we have:

$$u_1\sigma_1(y)\sigma_1(x) + \cdots + u_r\sigma_r(y)\sigma_r(x) = 0 \tag{3.3}$$

for all $x \in F$. Multiplying (3.2) by $\sigma_r(y)$, we have:

$$u_1\sigma_r(y)\sigma_1(x) + \cdots + u_r\sigma_r(y)\sigma_r(x) = 0 \tag{3.4}$$

for all $x \in F$. Subtracting (3.3) and (3.4), we have:

$$u_1(\sigma_1(y) - \sigma_r(y))\sigma_1(x) + \cdots + u_{r-1}(\sigma_{r-1}(y) - \sigma_r(y))\sigma_{r-1}(x) = 0$$

which is a shorter dependence relation, a contradiction. □

THEOREM 5.3.6. *Let F/k be a finite dimensional extension of fields. Then the order of the group of automorphisms $\mathrm{Aut}_k(F)$ is less than or equal to $\dim_k(F)$.*

PROOF. If $\mathrm{Aut}_k(F) = \langle 1 \rangle$, then there is nothing to prove. Let $r = \dim_k(F)$. For sake of contradiction assume $\mathrm{Aut}_k(F)$ contains a set of $r+1$ distinct automorphisms, which we enumerate: $\sigma_0, \ldots, \sigma_r$. Let $v_1, \ldots, v_r$ be a basis for F as a k-vector space. By Theorem 4.2.4, the $r+1$ vectors

$$\begin{aligned} x_0 &= \big(\sigma_0(v_1), \sigma_0(v_2), \ldots, \sigma_0(v_r)\big), \\ x_1 &= \big(\sigma_1(v_1), \sigma_1(v_2), \ldots, \sigma_1(v_r)\big), \\ &\vdots \\ x_r &= \big(\sigma_r(v_1), \sigma_r(v_2), \ldots, \sigma_r(v_r)\big), \end{aligned}$$

in F^r are linearly dependent over F. Hence there exists a nonzero vector $(c_0, c_1, \ldots, c_r)$ in F^{r+1} such that

$$c_0\sigma_0(v_j) + c_1\sigma_1(v_j) + \cdots + c_r\sigma_r(v_j) = 0 \tag{3.5}$$

for $j = 1, \ldots, r$. Let u be an arbitrary element of F. In terms of the k-basis, u has a representation $u = a_1v_1 + \cdots + a_rv_r$ for unique $a_1, \ldots, a_r$ in k. For each σ_i we have:

$$\sigma_i(u) = a_1\sigma_i(v_1) + \cdots + a_r\sigma_i(v_r). \tag{3.6}$$

Consider:

$$\begin{aligned}\sum_{i=0}^{r} c_i\sigma_i(u) &= \sum_{i=0}^{r} c_i\left(a_1\sigma_i(v_1)+\cdots+a_r\sigma_i(v_r)\right)\\ &= \sum_{j=1}^{r} a_j\left(\sum_{i=0}^{r} c_i\sigma_i(v_j)\right)\\ &= 0\end{aligned} \tag{3.7}$$

where the last equation follows from (3.5). Since u was arbitrary, (3.7) is a contradiction to Theorem 5.3.5. □

EXAMPLE 5.3.7. Let $F=\mathbb{F}_q$ be a finite field with order q and $\text{char}(F)=p$. If $k=\mathbb{F}_p$ is the prime subfield and $\dim_k(F)=n$, then $q=p^n$. By Exercise 3.2.31, the Frobenius homomorphism $\sigma: F\to F$ defined by $\sigma(x)=x^p$ is a homomorphism. Since F is a finite field, σ is necessarily one-to-one and onto (Exercises 3.2.28 and 1.1.11). Let α be a generator for the group of units of F (Corollary 3.6.10). Then in F^*, the order of α is $|\alpha|=p^n-1$. Therefore $\alpha^{p^n}=\alpha$ and if $1<i<p^n$, then $\alpha^i\neq\alpha$. It follows from $\sigma(\alpha)=\alpha^p\neq\alpha$, $\sigma^2(\alpha)=\sigma(\alpha^p)=(\alpha^p)^p=\alpha^{p^2}\neq\alpha,\ldots,\sigma^i(\alpha)=\alpha^{p^i}\neq\alpha,\ldots,\sigma^n(\alpha)=\alpha^{p^n}=\alpha$ that σ has order n in $\text{Aut}(F)$. By Theorem 5.3.6, $\text{Aut}(F)$ is cyclic and the Frobenius homomorphism σ is a generator.

If G is a group and H is a subgroup, the index of H in G is denoted $[G:H]$. The order of G is $[G:1]$.

THEOREM 5.3.8. *Let F/k be an extension of fields, G a finite subgroup of* $\text{Aut}_k(F)$*, and $K=F^G$. Then F/K is finite dimensional and* $\dim_K(F)\leq[G:1]$.

PROOF. Assume $[G:1]=n$ and $G=\{\sigma_1,\ldots,\sigma_n\}$. For sake of contradiction, assume the statement of the theorem is false. By Exercise 4.2.22, there exists a subset $\{v_0,\ldots,v_n\}\subseteq F$ which is linearly independent over K. By Theorem 4.2.4, the $n+1$ vectors

$$\begin{aligned}x_0&=\big(\sigma_1(v_0),\sigma_2(v_0),\ldots,\sigma_n(v_0)\big),\\ x_1&=\big(\sigma_1(v_1),\sigma_2(v_1),\ldots,\sigma_n(v_1)\big),\\ &\vdots\\ x_n&=\big(\sigma_1(v_n),\sigma_2(v_n),\ldots,\sigma_n(v_n)\big)\end{aligned}$$

in F^n are linearly dependent over F. Let V be the subspace of F^n spanned by $X=\{x_0,x_1,\ldots,x_n\}$. Then $\dim_F(V)\leq n$ so a linearly independent subset of X has cardinality at most n. By Corollary 4.2.6, there is a linearly independent subset of X that is a spanning set for V. If necessary, reorder the vectors in X such that x_0 is in the linear span of $\{x_1,\ldots,x_n\}$. If c_0 is an arbitrary element of F, then there exist n elements $c_1,\ldots,c_n$ in F such that $0=c_0x_0+c_1x_1+\cdots+c_nx_n$. This is equivalent to

$$0=\sum_{i=0}^{n} c_i\sigma_j(v_i) \tag{3.8}$$

for $j=1,\ldots,n$. For each $i=0,\ldots,n$, consider

$$a_i=\sigma_1(c_i)+\cdots+\sigma_n(c_i).$$

By Theorem 5.3.5, $\sigma_1,\ldots,\sigma_n$ are linearly independent so we can find c_0 in F such that $a_0\neq 0$. By the comment above, we can pick $c_1,\ldots,c_n$ so that (3.8) holds for $j=1,\ldots,n$.

Since G is a group,

$$\begin{aligned}\sigma_j(a_i) &= \sigma_j\sigma_1(c_i) + \cdots + \sigma_j\sigma_n(c_i)\\ &= \sigma_1(c_i) + \cdots + \sigma_n(c_i)\\ &= a_i\end{aligned}$$

implies $a_i \in K = F^G$, for $i = 0, 1, \ldots, n$. Consider

$$\begin{aligned}\sum_{i=0}^{n} a_i v_i &= \sum_{i=0}^{n}\left(\sum_{j=1}^{n}\sigma_j(c_i)\right)v_i\\ &= \sum_{i=0}^{n}\left(\sum_{j=1}^{n}\sigma_j(c_i)\sigma_j(\sigma_j^{-1}(v_i))\right)\\ &= \sum_{j=1}^{n}\sigma_j\left(\sum_{i=0}^{n}c_i\sigma_j^{-1}(v_i)\right)\\ &= 0\end{aligned} \tag{3.9}$$

where the last 0 is from (3.8). The left hand side of (3.9) is a nontrivial K-linear combination of $v_0, v_1, \ldots, v_n$. This is a contradiction. □

3.2. Galois Extensions. In this section useful necessary and sufficient conditions for an extension of fields F/k to be a Galois extension are derived. As an application, in Corollary 5.3.17 we prove the important result that any finite separable extension can be embedded as an intermediate field of a Galois extension.

DEFINITION 5.3.9. Let F/k be an extension of fields and G a finite subgroup of $\mathrm{Aut}_k(F)$. If $k = F^G$, then we say F/k is a *Galois* extension with Galois group G. We also say F is a G-Galois extension of k.

PROPOSITION 5.3.10. *If F is a G-Galois extension of k, then* $\dim_k(F) = [G:1]$ *and* $G = \mathrm{Aut}_k(F)$.

PROOF. This follows directly from Theorems 5.3.6 and 5.3.8. □

PROPOSITION 5.3.11. *Let F be a G-Galois extension of k and $\alpha \in F$. The subgroup of G fixing α is denoted G_α (Definition 2.4.9). If $G_\alpha = \langle 1\rangle$, then $F = k(\alpha)$.*

PROOF. Let $f = \min.\mathrm{poly}_k(\alpha)$. The orbit of α under the group G is $R = \{\sigma(\alpha) \mid \sigma \in G\}$. If $\sigma, \tau \in G$ and $\sigma(\alpha) = \tau(\alpha)$, then $\sigma^{-1}\tau \in G_\alpha = \langle 1\rangle$. Therefore, $|R| = [G:1]$. By Proposition 5.3.2, every element of R is a root of f. So $\deg f \geq [G:1]$. By Theorem 5.1.4, $\dim_k k(\alpha) = \deg f$. By Proposition 5.3.10, all of the numbers in the string of inequalities:

$$[G:1] \leq \deg f = \dim_k k(\alpha) \leq \dim_k(F)$$

are equal. Hence $k(\alpha) = F$. □

PROPOSITION 5.3.12. *Let F/k be a finite dimensional extension of fields and $\sigma_1, \ldots, \sigma_n$ a finite set of distinct automorphisms in* $\mathrm{Aut}_k(F)$. *If* $\dim_k(F) = n$, *then F/k is Galois with group* $\mathrm{Aut}_k(F) = \{\sigma_1, \ldots, \sigma_n\}$.

PROOF. We have $\{\sigma_1, \ldots, \sigma_n\} \subseteq \mathrm{Aut}_k(F)$, hence $n \leq [\mathrm{Aut}_k(F):1]$. By Theorem 5.3.6, $n = \dim_k(F) \geq [\mathrm{Aut}_k(F):1] \geq n$. Therefore, $\mathrm{Aut}_k(F) = \{\sigma_1, \ldots, \sigma_n\}$. In particular, this proves the set $\{\sigma_1, \ldots, \sigma_n\}$ is a group. For notational simplicity, let $G = \mathrm{Aut}_k(F)$ and $K = F^G$. By Theorem 5.3.8, $\dim_K(F) = n$. By Exercise 5.1.23 applied to the tower of fields: $k \subseteq K \subseteq F$, we conclude that $k = K = F^G$. □

EXAMPLE 5.3.13. Let $F = \mathbb{F}_q$ be a finite field with characteristic $\text{char}(F) = p$ and order q. If $k = \mathbb{F}_p$ is the prime subfield and $\dim_k(F) = n$, then $q = p^n$. By Example 5.3.7, $\text{Aut}_k(F) = \langle\sigma\rangle$ is cyclic of order n where σ is the Frobenius homomorphism defined by $\sigma(x) = x^p$. By Proposition 5.3.12, $\mathbb{F}_q/\mathbb{F}_p$ is Galois with cyclic Galois group.

DEFINITION 5.3.14. Let F/k be an algebraic extension of fields. We say F/k is a *normal* extension if every irreducible polynomial over k that has a root in F actually splits over F.

Theorem 5.3.15 provides very useful necessary and sufficient conditions for an extension of fields to be Galois.

THEOREM 5.3.15. *Let F/k be a finite dimensional extension of fields. The following are equivalent.*

(1) F/k is a Galois extension.
(2) F/k is normal and separable.
(3) F is the splitting field over k of a separable polynomial in $k[x]$.

PROOF. (1) implies (2): Suppose F/k is Galois with group $G = \{\sigma_1, \ldots, \sigma_n\}$. We prove F/k is normal and separable. Let $f(x) \in k[x]$ be an irreducible polynomial and suppose $u \in F$ is a root of f. Look at the orbit of α under G: $R = \{\sigma_1(u), \ldots, \sigma_n(u)\}$. Suppose R has r elements which we enumerate: $R = \{u_1, \ldots, u_r\}$. Then G acts as a group of permutations of R. The polynomial $g(x) = (x-u_1)(x-u_2)\cdots(x-u_r)$ is in $F[x]$ and is fixed by every element of G. Since $k = F^G$, we have $g(x) \in k[x]$. Now $u \in R$, so $g(u) = 0$. Since $f(x)$ is the irreducible polynomial of u, by Theorem 5.1.4 we have $f \mid g$. This proves f splits over F. Since g is separable, so is f. We have proved that F/k is normal. Let v be an arbitrary element of F. Then by the previous argument, $\text{min.poly}_k(v)$ is separable. This proves F/k is separable.

(2) implies (1): By Theorem 5.2.14, $F = k(\alpha)$ for some $\alpha \in F$. If $f = \text{Irr.poly}_k(\alpha)$, then f is separable and splits over F. If $n = \deg(f)$, then by Theorem 5.1.4, $n = \dim_k(F)$. Let $\alpha_1, \ldots, \alpha_n$ be the distinct roots of f in F. Then for each i we have $f = \text{Irr.poly}_k(\alpha_i)$. Since $k(\alpha_i$ is an intermediate field of F/k and $\dim_k k(\alpha_i) = \dim_k F$, we have $F = k(\alpha_i)$. By Corollary 5.1.7 there is a k-automorphism $\sigma_i : F \to F$ such that $\sigma_i(\alpha) = \alpha_i$. By Proposition 5.3.2 (2), $\sigma_1, \ldots, \sigma_n$ are distinct elements of $\text{Aut}_k(F)$. By Proposition 5.3.12, F/k is Galois.

(2) implies (3): By Theorem 5.2.14, The Primitive Element Theorem, $F = k(\alpha)$ for some $\alpha \in F$. If $f = \text{Irr.poly}_k(\alpha)$, then f is separable and splits in F.

(3) implies (1): Suppose $f \in k[x]$ is separable and F is the splitting field for f over k. Let $n = \dim_k(F)$. If $n = 1$, then $F = k$, so F/k is Galois with group $\langle 1\rangle$. Inductively, assume $n > 1$ and that (3) implies (1) for any extension of fields of dimension less than n. Let $G = \text{Aut}_k(F)$. To finish the proof, we show $F^G = k$. Let g be a monic irreducible factor of the polynomial f and assume $\deg g = d > 1$. Since g is separable and splits in F, there are d distinct roots $\alpha_1, \ldots, \alpha_d$ in F and $g = (x-\alpha_1)\cdots(x-\alpha_d)$. Now $k(\alpha_1)$ is an intermediate field of F/k and F is a splitting field of the separable polynomial f over $k(\alpha_1)$. By the induction hypothesis, we can assume $F/k(\alpha_1)$ is a Galois extension with group H and $[H : 1] = \dim_{k(\alpha_1)}(F)$. By Corollary 5.1.7, for each i, there is a k-algebra isomorphism $\sigma_i : k(\alpha_1) \to k(\alpha_i)$. By Lemma 5.2.7, each σ_i extends to an automorphism also denoted σ_i, in $G = \text{Aut}_k(F)$. Let θ be an arbitrary element of F^G. Since H is a subgroup of $G = \text{Aut}_k(F)$, $\theta \in F^H = k(\alpha_1)$. By Theorem 5.1.4(3) there are $c_0, c_1, \ldots, c_{d-1}$ in k such that

$$\theta = c_0 + c_1\alpha_1 + \cdots + c_{d-1}\alpha_1^{d-1}. \tag{3.10}$$

Applying σ_i to (3.10) we have

$$\theta = c_0 + c_1\alpha_i + \cdots + c_{d-1}\alpha_i^{d-1} \tag{3.11}$$

since θ is fixed by G. Let $p(x) = (c_0 - \theta) + c_1 x + \cdots + c_{d-1}x^{d-1} \in k(\alpha_i)[x]$. Then in F, there are d distinct roots $\alpha_1, \ldots, \alpha_d$ of $p(x)$. Since $\deg p(x) \leq d-1$, we must have $p = 0$. In particular, $\theta = c_0$ is in k. □

COROLLARY 5.3.16. *Let k be a field, f an irreducible separable polynomial in $k[x]$, and F a splitting field for f over k. If $n = \deg(f)$, then the following are true:*

(1) F/k is a Galois extension with group $G = \mathrm{Aut}_k(F)$.
(2) G acts as a group of permutations of the roots $\alpha_1, \ldots, \alpha_n$ of f.
(3) G is isomorphic to a subgroup of S_n, the symmetric group on n letters.

PROOF. By Theorem 5.3.15, F/k is Galois. By Exercise 5.3.29, G acts on the roots of f. There is a homomorphism $\theta : G \to S_n$. Since $F = k(\alpha_1, \ldots, \alpha_n)$, if two automorphisms define the same permutation of $\alpha_1, \ldots, \alpha_n$, they define the same automorphism of F. This proves θ is one-to-one. □

COROLLARY 5.3.17. *(Embedding Theorem for Fields) Let F/k be a finite dimensional extension of fields. If F/k is separable, then there exists a finite dimensional Galois extension K/k which contains F as an intermediate field.*

PROOF. Pick a finite set of separable elements $u_1, \ldots, u_n$ that generate F/k. For each i, if $f_i = \mathrm{Irr.poly}_k(u_i)$, then f_i is separable over k. Let K be the splitting field for $f_1 \cdots f_n$ over k. So K contains a generating set for F, hence F is an intermediate field of K/k. By Theorem 5.3.15, K/k is a Galois extension. □

3.3. The Fundamental Theorem of Galois Theory. In this section, we prove the Fundamental Theorem of Galois Theory. To illustrate the theorem, nontrivial examples are given for which the Galois group is completely determined.

THEOREM 5.3.18. *(The Fundamental Theorem of Galois Theory) Let F/k be a Galois extension of fields with finite group G. There is a one-to-one order inverting correspondence between the subgroups H of G and the intermediate fields E of F/k. A subgroup H corresponds to the fixed field F^H. An intermediate field E corresponds to the subgroup of G fixing E, G_E. If E is an intermediate field of F/k, then*

(1) $\dim_E(F) = [G_E : 1]$, $\dim_k(E) = [G : G_E]$, $G_E = \mathrm{Aut}_E(F)$,
(2) F/E is a Galois extension with group G_E, and
(3) E/k is a Galois extension if and only if G_E is a normal subgroup of G and in this case, $G/G_E \cong \mathrm{Aut}_k(E)$.

PROOF. By Proposition 5.3.1 there are functions

$$\{H \mid H \text{ is a subgroup of } G\} \underset{\lambda}{\overset{\rho}{\rightleftarrows}} \{E \mid E \text{ is an intermediate field of } F/k\}$$

defined by $\rho(H) = F^H$ and $\lambda(E) = G_E$. It is clear that if $H_1 \subseteq H_2$, then $\rho(H_1) \supseteq \rho(H_2)$. Likewise, if $E_1 \subseteq E_2$, then $\lambda(E_1) \supseteq \lambda(E_2)$. Suppose A and B are two subgroups of G such that $F^A = F^B$. Let $E = F^A = F^B$. Proposition 5.3.10 says $\dim_E(F) = [A : 1] = [B : 1]$. If there exists $\sigma \in B - A$, then we get a contradiction to Theorem 5.3.6. So $B \subseteq A$. Similarly, $A \subseteq B$. This shows ρ is one-to-one. Let E be an intermediate field of F/k. Since F/k is Galois, by Theorem 5.3.15, F is the splitting field of a separable polynomial f in $k[x]$.

Then f is a separable polynomial in $E[x]$ and F is the splitting field of f over E. Since F/E is finite dimensional, Theorem 5.3.15 implies F/E is Galois. By Proposition 5.3.2, $G_E = \mathrm{Aut}_E(F)$ is the subgroup of $\mathrm{Aut}_k(F)$ fixing E and $\dim_E(F) = [G_E : 1]$. This implies $E = \rho\lambda(E)$, so ρ is a one-to-one correspondence. By Lagrange's Theorem (Corollary 2.2.12), $[G : 1] = [G : G_E][G_E : 1]$. By Exercise 5.1.23 $\dim_k(F) = \dim_k(E)\dim_E(F)$. This says $\dim_k(E) = [G : G_E]$. We have proved (1) and (2).

The rest of the proof is devoted to proving (3). Assume E/k is Galois. We prove that $G_E = \mathrm{Aut}_E(F)$ is a normal subgroup of $G = \mathrm{Aut}_k(F)$ and $\mathrm{Aut}_k(E)$ is isomorphic to the quotient G/G_E. First we show that there is a homomorphism of groups:

$$G = \mathrm{Aut}_k(F) \xrightarrow{h} \mathrm{Aut}_k(E)$$

defined by $\phi \mapsto \phi|_E$. The binary operation in both groups is composition of functions, so it suffices to show that if $\phi \in G$, then $\phi(E) = E$. By Theorem 5.2.14, $E = k(\xi)$ is a simple extension. Say $g(x) = \min.\mathrm{poly}_k(\xi)$ and $\deg g = m$. Since E/k is normal, g splits over E and has m distinct roots in E, call them $\xi_1, \ldots, \xi_m$. Given $\phi \in \mathrm{Aut}_k(F)$, $\phi(\xi) = \xi_j$ for some j, by Proposition 5.3.2. Therefore, $\phi(E) = \phi(k(\xi)) \subseteq E$. Since ϕ is one-to-one, $\phi(E) = E$ by Theorem 4.6.12. From this is follows that $\phi|_E$ is an automorphism of E, and h is a homomorphism of groups. The kernel of h is G_E, the set of all $\phi \in G$ fixing E. Therefore, $G_E = \mathrm{Aut}_E(F)$ is a normal subgroup of $G = \mathrm{Aut}_k(F)$. To show that $\mathrm{Aut}_k(E)$ is isomorphic to the quotient G/G_E, it suffices to show h is onto (Theorem 2.3.12). We are given that E/k is a Galois extension. This and (1) yield $[\mathrm{Aut}_k(E) : 1] = \dim_k(E) = [G : G_E]$. Theorem 2.3.12 and Lagrange's Theorem (Corollary 2.2.12) yield: $[G : G_E] = [\mathrm{im}\, h : 1]$. Therefore, $[\mathrm{im}\, h : 1] = [\mathrm{Aut}_k(E) : 1]$. Since the groups are finite, h is onto.

Conversely, assume $G_E = \mathrm{Aut}_E(F)$ is a normal subgroup of G and prove E/k is Galois. First we show that there is a homomorphism of groups

$$G = \mathrm{Aut}_k(F) \xrightarrow{h} \mathrm{Aut}_k(E)$$

defined by $\psi \mapsto \psi|_E$. To show that h is well defined, we use the fact that $\psi^{-1}\mathrm{Aut}_E(F)\psi = \mathrm{Aut}_E(F)$ (Lemma 2.3.4). Let $\phi \in \mathrm{Aut}_E(F)$. Then $\psi^{-1}\phi\psi = \phi_1 \in \mathrm{Aut}_E(F)$. Let y be an arbitrary element of E. Then $\psi^{-1}\phi\psi(y) = \phi_1(y) = y$. Therefore, $\phi\psi(y) = \psi(y)$. This shows $\psi(y)$ is fixed by each ϕ in $\mathrm{Aut}_E(F)$. By (2), this means $\psi(y) \in E$, hence h is well defined. The kernel of h is $G_E = \mathrm{Aut}_E(F)$, the subgroup of $\mathrm{Aut}_k(F)$ fixing E. By Theorem 2.3.11, the diagram

$$\begin{array}{ccc} \mathrm{Aut}_k(F) & \xrightarrow{\quad h \quad} & \mathrm{Aut}_k(E) \\ & {\scriptstyle\eta}\searrow \quad \nearrow{\scriptstyle\bar{h}} & \\ & \mathrm{Aut}_k(F)/\mathrm{Aut}_E(F) & \end{array}$$

commutes and $\bar{h}$ is one-to-one. From (1) and Lagrange's Theorem,

$$\begin{aligned} \dim_k(E) &= \dim_k(F)/\dim_E(F) \\ &= [\mathrm{Aut}_k(F) : 1]/[\mathrm{Aut}_E(F) : 1] \\ &= [\mathrm{im}(h) : 1] \\ &\leq [\mathrm{Aut}_k(E) : 1] \end{aligned}$$

By Proposition 5.3.12, E/k is Galois. □

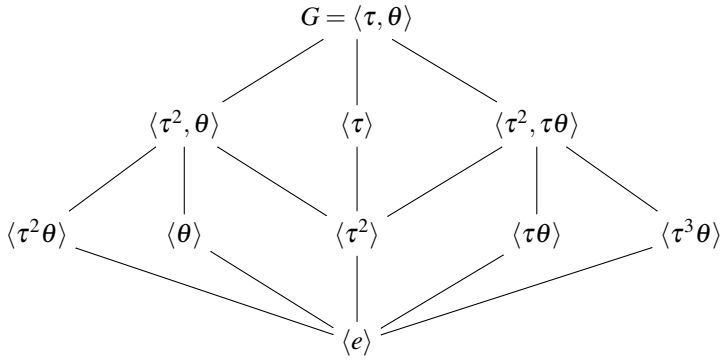
EXAMPLE 5.3.19. This is an example of a Galois extension of $\mathbb{Q}$ with Galois group the full symmetric group S_p. Let p be a prime number and $f \in \mathbb{Q}[x]$ an irreducible polynomial

of degree p such that f has exactly two nonreal roots. In this example we show that the Galois group of f is isomorphic to S_p, the symmetric group on p letters. Let F be the splitting field for f in $\mathbb{C}$. By Theorem 5.3.15, F is Galois over $\mathbb{Q}$. Let a and b be the nonreal roots of f. If $p = 2$, then $F = \mathbb{Q}(a)$ has degree two over $\mathbb{Q}$ and $\mathrm{Aut}_{\mathbb{Q}}(F)$ has order two hence is isomorphic to S_2. Assume $p > 2$ and let c be a real root of f. Then $\dim_{\mathbb{Q}} \mathbb{Q}(c) = p$ and by Theorem 5.3.18, p divides the order of $\mathrm{Aut}_{\mathbb{Q}}(F)$. By Cauchy's Theorem (Corollary 2.4.14), $\mathrm{Aut}_{\mathbb{Q}}(F)$ contains an element σ of order p. By Corollary 5.3.16, we know that $\mathrm{Aut}_{\mathbb{Q}}(F)$ is a group of permutations of the roots of f. By Corollary 2.6.4 we know that σ is a p-cycle and can be written in the form $\sigma = (s_1 s_2 \cdots s_p)$. For some i and j we have $a = s_i$ and $b = s_j$. Then $\sigma^{j-i}(s_i) = s_j$. Therefore, we can write σ^{j-i} in the cycle form $(abt_3 \cdots t_p)$. Let χ be the automorphism of $\mathbb{C}$ defined by complex conjugation (Example 5.3.4). Then χ maps F to F. Also, $\chi(a) = b$ and χ fixes every real root of f. So χ corresponds to the transposition $\chi = (ab)$. By Exercise 2.6.16, the group S_p is generated by the transposition (12) and the p-cycle $(123 \cdots p)$. Therefore, $\mathrm{Aut}_{\mathbb{Q}}(F)$ is generated by χ and σ^{j-i}, hence is isomorphic to S_p.

EXAMPLE 5.3.20. In $\mathbb{Q}[x]$, let $f(x) = x^4 - 2$. Let u be the positive real number such that $u^4 = 2$ and let $i \in \mathbb{C}$ be a root of $x^2 + 1$. Then the four roots of $f(x)$ in $\mathbb{C}$ are

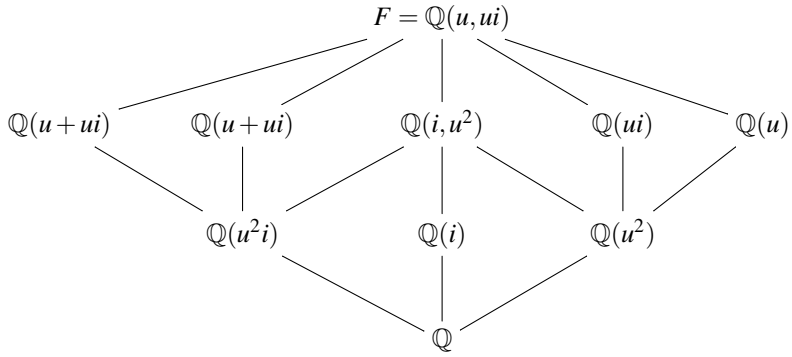
$$\{u, -u, ui, -ui\}. \tag{3.12}$$

Let $F = \mathbb{Q}(u, ui)$ be the splitting field of f over $\mathbb{Q}$. By Theorem 5.1.4, $(\mathbb{Q}(u) : \mathbb{Q}) = (\mathbb{Q}(ui) : \mathbb{Q}) = 4$. Since $u \in \mathbb{R}$ is real and $ui \notin \mathbb{R}$ is nonreal, we know $\mathbb{Q}(u) \neq \mathbb{Q}(ui)$. Over $\mathbb{Q}(u^2)$ we have the factorization $f = (x^2 - u^2)(x^2 + u^2)$ into irreducibles. The irreducible polynomial for ui over $\mathbb{Q}(u)$ is $x^2 + u^2$. The irreducible polynomial for u over $\mathbb{Q}(ui)$ is $x^2 - u^2$. Then $(F : \mathbb{Q}(u)) = (F : \mathbb{Q}(ui)) = 2$. By Corollary 5.1.7, there is an isomorphism $\sigma : \mathbb{Q}(u) \to \mathbb{Q}(ui)$ which is given by $\sigma(u) = ui$. By Lemma 5.2.7, σ can be extended to an isomorphism $F = \mathbb{Q}(u)(ui) \to \mathbb{Q}(ui)(u) = F$ which is defined by sending ui to one of u or $-u$. Let τ be the automorphism of F defined by $\tau(u) = ui$, $t(ui) = -u$. Let θ be the automorphism of F defined by $\tau(u) = ui$, $t(ui) = u$. By Theorem 5.3.15, F is Galois over $\mathbb{Q}$ with group $G = \mathrm{Aut}_{\mathbb{Q}}(F)$. By Exercise 5.3.29, we can view G as a subgroup of S_4. Using the ordering of the roots given in (3.12), the cycle representations of τ and θ are $\tau = (1324)$, $\theta = (13)(24)$. We can now compute the elements of G: $\langle e, \tau = (1324), \tau^2 = (12)(34), \tau^3 = (1423), \theta = (13)(24), \tau\theta = (12), \tau^2\theta = (14)(23), \tau^3\theta = (34)\rangle$. Therefore, G is isomorphic to the dihedral group D_4 (Example 2.1.16). The subgroup lattice of G was computed in Example 2.3.37:



By Example 2.3.32, the center of G is $\langle \tau^2 \rangle$ which is normal. The three subgroups of order four are normal. The other four subgroups of order two, $\langle \theta \rangle$, $\langle \tau^2 \rangle$, $\langle \tau\theta \rangle$, and

$\langle \tau^3\theta \rangle$, are not normal. Notice that $\tau^3\theta$ is complex conjugation. The reader should verify that $F^{\langle \tau^2 \rangle} = \mathbb{Q}(i,u^2)$, $F^{\langle \tau\theta \rangle} = \mathbb{Q}(ui)$, $F^{\langle \tau^3\theta \rangle} = \mathbb{Q}(u)$, $F^{\langle \tau^2,\theta \rangle} = \mathbb{Q}(u^2 i)$, $F^{\langle \tau \rangle} = \mathbb{Q}(i)$, $F^{\langle \tau^2,\tau\theta \rangle} = \mathbb{Q}(u^2)$, $F^{\langle \tau^2\theta \rangle} = \mathbb{Q}(u+ui)$, and $F^{\langle \theta \rangle} = \mathbb{Q}(u+ui)$. The subfield lattice of F is



Notice that $\mathbb{Q}(u)$ is Galois over $\mathbb{Q}(u^2)$, $\mathbb{Q}(u^2)$ is Galois over $\mathbb{Q}$, but $\mathbb{Q}(u)$ is not Galois over $\mathbb{Q}$. This example shows that the property of being Galois is not transitive. In other words, Galois over Galois is not Galois. The analogous statement for groups is also true. Namely, normal over normal is not normal.

3.4. Exercises.

EXERCISE 5.3.21. Show that the group of automorphisms of a prime field is trivial. In other words, prove: $\mathrm{Aut}(\mathbb{Q}) = \langle 1 \rangle$ and $\mathrm{Aut}(\mathbb{Z}_p) = \langle 1 \rangle$. (Hint: Exercise 3.2.48.)

EXERCISE 5.3.22. Let F be a field, k the prime field of F, and σ an automorphism of F. Show that $\sigma(a) = a$ for every $a \in k$.

EXERCISE 5.3.23. This exercise outlines a proof that $\mathrm{Aut}(\mathbb{R}) = \langle 1 \rangle$. In the following, assume a,b,c are real numbers and r,s are rational numbers. For this exercise you can assume that if $a < b$, then there exists a rational number r such that $a < r < b$. Let σ be an automorphism of $\mathbb{R}$. Prove:

(1) $\sigma(a^2) = \sigma(a)^2$.
(2) If $b > 0$, then $\sigma(b) > 0$.
(3) If $r < c < s$, then $r < \sigma(c) < s$.
(4) For every $c \in \mathbb{R}$, $\sigma(c) = c$.

EXERCISE 5.3.24. Let $f(x) = x^3 + 3x + 3$. Show that f is irreducible in $\mathbb{Q}[x]$ and f has exactly one real root and two nonreal roots. Let $\alpha \in \mathbb{R}$ be the real root and β_1, β_2 be the nonreal roots of $f(x)$. Show that $\mathbb{Q}[\alpha,\beta_1]$ is the splitting field for f over $\mathbb{Q}$ and $\dim_{\mathbb{Q}} \mathbb{Q}[\alpha,\beta_1] = 6$. Show that $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}[\alpha]) = \langle 1 \rangle$. Show that $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}[\alpha,\beta_1])$ is isomorphic to S_3, the group of permutations of $\{\alpha,\beta_1,\beta_2\}$.

EXERCISE 5.3.25. Prove the following for $f = x^3 + x - 1$.

(1) f is irreducible in $\mathbb{Q}[x]$.
(2) If $F = \mathbb{Q}[x]/(f)$ and σ is an automorphism of F, then σ is the identity function.
(3) In $\mathbb{R}[x]$, f factors into a product of a linear polynomial and an irreducible quadratic.
(4) If F is the splitting field of f over $\mathbb{Q}$, then the Galois group $\mathrm{Aut}_{\mathbb{Q}}(F)$ is a nonabelian group of order six.

EXERCISE 5.3.26. Let F be the splitting field of $f = x^3 - 5$ over $\mathbb{Q}$.

(1) Show that the Galois group $\mathrm{Aut}_{\mathbb{Q}}(F)$ is a nonabelian group of order six.
(2) Find all intermediate fields K between $\mathbb{Q}$ and F.
(3) Prove or give a counterexample: Each intermediate field K is a Galois extension of $\mathbb{Q}$.

EXERCISE 5.3.27. Let F be the splitting field of $f = (x^2-2)(x^2-3)$ over $\mathbb{Q}$.

(1) Show that the Galois group $\mathrm{Aut}_{\mathbb{Q}}(F)$ is a noncyclic abelian group of order four.
(2) Find all intermediate fields K between $\mathbb{Q}$ and F.
(3) Prove or give a counterexample: Each intermediate field K is a Galois extension of $\mathbb{Q}$.

EXERCISE 5.3.28. Consider the polynomial $f = x^4 + p^2$ in $\mathbb{Q}[x]$, where p is a prime number. Determine the following.

(1) The splitting field of f over $\mathbb{Q}$. Call this field K.
(2) The Galois group of f over $\mathbb{Q}$.
(3) The lattice of intermediate fields of $K/\mathbb{Q}$. Determine which intermediate fields are normal over $\mathbb{Q}$.

EXERCISE 5.3.29. Let $f \in k[x]$ be an irreducible separable polynomial of degree n over the field k. Let F/k be the splitting field for f over k and let $G = \mathrm{Aut}_k(F)$ be the Galois group. We call G the *Galois group* of f. Prove the following.

(1) G acts transitively on the roots of f. That is, given two roots α, β of f, there is $\sigma \in G$ such that $\sigma(\alpha) = \beta$. (Hint: apply Corollary 5.1.7 and Lemma 5.2.7.)
(2) n divides $[G:1]$.

EXERCISE 5.3.30. Let F be a field and $f(x)$ a polynomial in $F[x]$ such that $f'(x) = 0$. That is, the derivative of $f(x)$ is the zero polynomial.

(1) If F has characteristic 0, show that $f(x) = \alpha$, for some $\alpha \in F$.
(2) If F has characteristic $p > 0$, show that there exists $g(x) \in F[x]$ such that $f(x) = g(x^p)$.

EXERCISE 5.3.31. Let F/k be an extension of fields. Let $\alpha \in F$. Prove that $F(\alpha)$ is a separable extension of k if and only if α is separable over k.

EXERCISE 5.3.32. Let F/k be a separable extension of fields such that $\dim_k(F) = 2$. Prove that F/k is a Galois extension.

EXERCISE 5.3.33. Let F be a field, $\phi \in \mathrm{Aut}(F)$ and $k = F^{\langle \phi \rangle}$. Let $f \in F[x]$ be a polynomial satisfying:

(1) f is monic,
(2) f splits in $F[x]$,
(3) f has no repeated root, and
(4) if $\alpha \in F$ and $f(\alpha) = 0$, then $f(\phi(\alpha)) = 0$.

Show that $f \in k[x]$.

EXERCISE 5.3.34. Let F/k be an extension of fields where $\mathrm{char}\, k = p > 0$. Let $\alpha \in F$. Prove that α is separable over k if and only if $k(\alpha) = k(\alpha^p)$.

EXERCISE 5.3.35. Determine the group of automorphisms $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\sqrt{2}))$.

EXERCISE 5.3.36. Let p be a prime number and $\zeta = e^{2\pi i/p}$ a primitive pth root of unity in $\mathbb{C}$. Show that the group of automorphisms $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta))$ is isomorphic to the group of units in the ring $\mathbb{Z}/p$, hence is a cyclic group of order $p-1$.

EXERCISE 5.3.37. Let F be the splitting field for x^3-2 over $\mathbb{Q}$ (see Exercise 5.2.16). Show that the group of automorphisms $\mathrm{Aut}_{\mathbb{Q}}(F)$ is isomorphic to the symmetric group S_3.

EXERCISE 5.3.38. Let F/k be a Galois extension of fields with finite group G. Let α be an arbitrary element of F, and set

$$g = \prod_{\sigma \in G} (x - \sigma(\alpha)).$$

Show that $g \in k[x]$ and the only irreducible factor of g in $k[x]$ is $\mathrm{Irr.poly}_k(\alpha)$.

EXERCISE 5.3.39. Determine the Galois group of the polynomial x^4+x^2-6 over $\mathbb{Q}$.

EXERCISE 5.3.40. Determine the smallest Galois extension $K/\mathbb{Q}$ containing $2^{1/2}+2^{1/3}$. Determine $\mathrm{Aut}_{\mathbb{Q}}(K)$.

EXERCISE 5.3.41. Determine the Galois group of the polynomial x^6-8 over each of these fields: $\mathbb{Q}$, $\mathbb{Q}(\sqrt{2})$, and $\mathbb{Q}(\zeta)$, where $\zeta = e^{2\pi i/3}$ is a primitive third root of 1 in $\mathbb{C}$.

EXERCISE 5.3.42. Determine the Galois group of the polynomial $\left(x^2-2\right)\left(x^3+2\right)$ over each of these fields: $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Q}(\sqrt{2})$, $\mathbb{Q}(\sqrt[3]{2})$, and $\mathbb{Q}(\zeta)$, where $\zeta = e^{2\pi i/6}$ is a primitive third root of -1 in $\mathbb{C}$.

EXERCISE 5.3.43. Let F denote the splitting field of x^8-1 over the field $\mathbb{Q}$ of rational numbers. Determine the lattice of subfields and show that the Galois group $\mathrm{Aut}_{\mathbb{Q}}(F)$ is a noncyclic group of order 4.

EXERCISE 5.3.44. Let k be a field of characteristic zero and f an irrreducible polynomial in $k[x]$. Let F/k be an extension of fields and assume f splits over F. Prove that if $\alpha \in F$ and $f(\alpha)=0$, then $f(\alpha+1) \neq 0$.

4. Separable Closure

Given an algebraic extension of fields F/k we construct the separable closure of k in F. This result is then applied to show that the property of being separable is transitive. As another application, we prove in Theorem 5.4.5 a characterization of perfect fields. As we saw in Example 5.3.20, the property of being Galois is not transitive. Nevertheless, we prove in Theorem 5.4.6 that the property of being Galois is preserved under a change of base field. As an application of Galois Theory, in Theorem 5.4.10 we give a proof of the Fundamental Theorem of Algebra.

4.1. The Existence of a Separable Closure.

LEMMA 5.4.1. *Let F/k be an extension of fields and assume* $\mathrm{char}\, k = p > 0$. *Let $u \in F$ and assume u is algebraic over k. There exists $n \geq 0$ such that u^{p^n} is separable over k.*

PROOF. If u is separable over k, then take $n=0$. Let $f = \mathrm{Irr.poly}_k(u)$ and use induction on the degree of f. Assume f is not separable and $d = \deg f > 1$. By Lemma 5.2.11, there exists $g \in k[x]$ such that $f(x) = g(x^p)$. Because f is irreducible, so is g. Therefore, $f(u) = g(u^p) = 0$, u^p is algebraic over k, and the degree of $\mathrm{Irr.poly}_k(u^p)$ is equal to d/p. By induction on d, there is some $n \geq 0$ such that $(u^p)^{p^n}$ is separable over k. □

THEOREM 5.4.2. *Let F/k be an algebraic extension of fields. If*

$$S = \{u \in F \mid u \text{ is separable over } k\},$$

then S is an intermediate field of F/k, and S/k is separable. The field S is called the separable closure of k in F.

PROOF. It is enough to show S is a field. Let α and β be elements of $S-k$. If $f = \text{Irr.poly}_k(\alpha)$, then f is separable and irreducible over k. Likewise, $g = \text{Irr.poly}_k(\beta)$ is separable and irreducible over k. By Theorem 5.3.15, if E is the splitting field over k of fg, then E/k is a separable extension of fields. Since $k(\alpha,\beta)$ is an intermediate field of E/k, it is itself a separable extension of k. Therefore, S contains $\alpha+\beta$, $\alpha-\beta$, $\alpha\beta$, α/β. It follows that S is a field. □

THEOREM 5.4.3. *(Separable over Separable is Separable) Let $k \subseteq F \subseteq K$ be a tower of algebraic field extensions. If F is separable over k and K is separable over F, then K is separable over k.*

PROOF. By Proposition 5.1.10 (4), K is algebraic over k. If $\text{char}\, k = 0$, then an algebraic extension is separable, so assume $\text{char}\, k = p > 0$. By Theorem 5.4.2, let S be the separable closure of k in K. Then $F \subseteq S \subseteq K$. It is enough to show $S = K$. Let $u \in K$. By Lemma 5.4.1, there exists $n \geq 0$ such that $\alpha = u^{p^n}$ is in S. Then u satisfies the polynomial $x^{p^n} - \alpha \in S[x]$ and in $K[x]$ we have the factorization $x^{p^n} - \alpha = (x-u)^{p^n}$. If $f = \text{Irr.poly}_S(u)$, then f divides $(x-u)^{p^n}$ in $K[x]$. If $g = \text{Irr.poly}_F(u)$, then g is separable and since f divides g in $S[x]$, we know that f has no multiple roots in K. So $f = x-u$ and $u \in S$. □

DEFINITION 5.4.4. A field k is said to be *perfect* if $\text{char}\, k = 0$, or $\text{char}\, k = p$ is a prime number and the Frobenius homomorphism $\theta : k \to k$ by $a \mapsto a^p$ is onto (see Exercise 3.2.31).

THEOREM 5.4.5. *Let k be a field. The following are equivalent.*

(1) k is a perfect field.
(2) Every irreducible polynomial in $k[x]$ is separable.
(3) Every algebraic extension of k is separable over k.

PROOF. If $\text{char}\, k = 0$, then this is immediate.

(2) is equivalent to (3): This is Exercise 5.4.11.

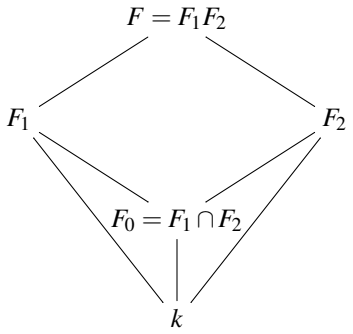
(3) implies (1): Assume k has positive characteristic p and every algebraic extension of k is separable. Let $\varphi : k \to k$ be the Frobenius homomorphism (Exercise 3.2.31). Let $\alpha \in k$. We show $\alpha = \varphi(u)$ for some $u \in k$. Consider the polynomial $x^p - \alpha$ in $k[x]$. Let F be an extension of k containing a root u of $x^p - \alpha$. In $F[x]$ we have the factorization $x^p - \alpha = (x-u)^p$. By assumption, F/k is separable, which implies this factorization occurs in $k[x]$. That is, $u \in k$ and $\alpha = \varphi(u)$.

(1) implies (3): Let F/k be an algebraic extension. Let $\alpha \in F - k$. Let $f \in k[x]$ be the irreducible polynomial of α over k. We show that $k(\alpha)$ is a separable extension of k. If $\text{char}\, k = 0$, it follows from Theorem 3.6.15 that f is separable and we are done. Assume $\text{char}\, k = p > 0$ and the Frobenius homomorphism $\varphi : k \to k$ is an automorphism of k. By Theorem 3.6.2, $\varphi(f) = g$ is an irreducible polynomial in $k[x]$ such that $\deg g = \deg f$. Since $g(\alpha^p) = (f(\alpha))^p = 0$, we see that $k(\alpha^p)$ is a field extension of k which is an intermediate field of $k(\alpha)/k$ such that $\dim_k(k(\alpha^p)) = \dim_k(k(\alpha))$. It follows that $k(\alpha^p) = k(\alpha)$, hence the Frobenius homomorphism is an automorphism $\varphi : k(\alpha) \to k(\alpha)$. For any $m > 0$, $\varphi^m(x) = x^{p^m}$. Since $k[\alpha] = k(\alpha)$, a typical element in $k(\alpha)$ can be represented in

the form $u = \sum_i a_i \alpha^i$ where $a_i \in k$. Therefore $\varphi^m(u) = \sum_i a_i^{p^m} (\alpha^{p^m})^i$ is in $k(\alpha^{p^m})$. This shows $k(\alpha^{p^m}) = k(\alpha)$ for all $m > 0$. By Theorem 5.4.2, let S be the separable closure of k in $k(\alpha)$. For some $n \geq 0$, $\alpha^{p^n} \in S$. Therefore $k(\alpha) = k(\alpha^{p^n}) \subseteq S$ so $k(\alpha)$ is a separable extension of k. □

4.2. A Change of Base Theorem for a Galois Extension. Theorem 5.4.6 is what is called a "change of base" theorem for a Galois extension. It says that if F_1/k is a Galois extension and F_2/k is a finite field extension, then $F = F_1F_2$ is a Galois extension of F_2. The base field is extended from k to F_2. This useful result also gives sufficient conditions such that the Galois group is preserved.

THEOREM 5.4.6. *Let K/k be a finite dimensional extension of fields. Let F_1 and F_2 be intermediate fields. Set $F = F_1F_2$ and $F_0 = F_1 \cap F_2$.*



(1) If F_1 is a Galois extension of k, then F is a Galois extension of F_2 and there is an isomorphism of groups $\mathrm{Aut}_{F_2}(F) \cong \mathrm{Aut}_{F_0}(F_1)$ defined by the assignment $\phi \mapsto \phi|_{F_1}$.

(2) If F_1 and F_2 are both Galois extensions of k, then F is a Galois extension of k. If $F_1 \cap F_2 = k$, then $\mathrm{Aut}_k(F) \cong \mathrm{Aut}_{F_1}(F) \times \mathrm{Aut}_{F_2}(F)$.

PROOF. (1): By Theorems 5.3.15 and 5.2.14, $F_1 = k(u)$ is a simple extension. Let $f = \mathrm{Irr.poly}_k(u)$. By Theorem 5.1.12, $F = F_2(u)$. Let $g = \mathrm{Irr.poly}_{F_2}(u)$. Theorem 5.1.4 implies g divides f. Then every root of g is in F, hence F is a splitting field for g. By Theorem 5.3.15, F/F_2 is a Galois extension. If $\phi \in \mathrm{Aut}_{F_2} F$, then ϕ is completely determined by the value of $\phi(u)$. But $\phi(u)$ is a root of f. Since F_1 is a splitting field for f, $\phi(F_1) \subseteq F_1$. Since ϕ fixes F_2 point-wise, ϕ fixes k point-wise. Therefore, $\theta : \mathrm{Aut}_{F_2}(F) \to \mathrm{Aut}_k(F_1)$ is a homomorphism of groups. If ϕ fixes F_1 point-wise, then $\phi(u) = u$ and ϕ is the identity function on F. This proves θ is one-to-one. Using θ, we identify $\mathrm{Aut}_{F_2}(F)$ with a subgroup of $\mathrm{Aut}_k F_1$. Let $E = F_1^{\mathrm{Aut}_{F_2}(F)}$. By Theorem 5.3.18, F_1/E is a Galois extension and $\dim_E(F_1) = |\mathrm{Aut}_{F_2}(F)| = \dim_{F_2}(F)$. Since $F_1 \subseteq F$, we have $E \subseteq F^{\mathrm{Aut}_{F_2}(F)} = F_2$. Since $\dim_{F_2}(F) = \dim_E(F_1)$, Exercise 5.1.23 implies that $\dim_E(F) = \dim_E(F_1)\dim_E(F_2)$. By Theorem 5.1.12 (5), we have $E = F_1 \cap F_2$, which completes the proof.

(2): This is Exercise 5.4.12. □

4.3. Examples. In this section we include some examples that did not seem to fit in elsewhere.

EXAMPLE 5.4.7. This is an example of a Galois extension of $\mathbb{Q}$ with abelian Galois group of order 8. Let a be a positive odd integer and $f = x^8 + a^4$. By Exercise 3.7.12, f is irreducible over $\mathbb{Q}$. Let ζ be the complex number $e^{2\pi i/16}$. Then $\zeta^8 = -1$. Let α be the

positive real number such that $\alpha^2 = a$. For any integer k, $f(\zeta^{2k+1}\alpha) = \zeta^8\zeta^{16k}\alpha^8 + a^4 = 0$. Therefore the eight roots of f in $\mathbb{C}$ are $S = \{\zeta^{2k+1}\alpha \mid 0 \le k \le 7\}$. By Theorem 5.1.4, the set $\{1, \zeta\alpha, \zeta^2\alpha^2, \ldots, \zeta^7\alpha^7\}$ is a basis for $\mathbb{Q}(\zeta\alpha)$ as a $\mathbb{Q}$-vector space. Since $(\zeta\alpha)^{2k+1} = \zeta^{2k+1}a^k\alpha$, we see that $S \subseteq \mathbb{Q}(\zeta\alpha)$. Hence $\mathbb{Q}(\zeta\alpha)$ is a splitting field for f. By Corollary 5.1.7 applied to $\zeta\alpha$ and $\zeta^3\alpha$, there is an automorphism $\tau \in \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha))$ such that $\tau(\zeta\alpha) = \zeta^3\alpha$. Since $\zeta^2\alpha^2 = \zeta^2 a$, it follows that $\zeta^2 \in \mathbb{Q}(\zeta\alpha)$. We have $\tau(\zeta^2) = \tau((\zeta\alpha)^2a^{-1}) = \tau(\zeta\alpha)^2a^{-1} = (\zeta^3\alpha)^2a^{-1} = (\zeta^6 a)a^{-1} = \zeta^6$. Using this it is now possible to compute the action of τ on S: $\tau(\zeta\alpha) = \zeta^3\alpha$, $\tau(\zeta^3\alpha) = -\zeta\alpha$, $\tau(-\zeta\alpha) = -\zeta^3\alpha$, $\tau(-\zeta^3\alpha) = \zeta\alpha$, $\tau(\zeta^5\alpha) = -\zeta^7\alpha$, $\tau(-\zeta^7\alpha) = -\zeta^5\alpha$, $\tau(-\zeta^5\alpha) = \zeta^7\alpha$, $\tau(\zeta^7\alpha) = \zeta^5\alpha$. So τ has two disjoint orbits, each of length four. Fix this ordering of the 8 elements of S:

$$S = \{\zeta\alpha, \zeta^3\alpha, -\zeta\alpha, -\zeta^3\alpha, \zeta^7\alpha, \zeta^5\alpha, -\zeta^7\alpha, -\zeta^5\alpha\}. \tag{4.1}$$

Then τ has the cycle representation $\tau = (1234)(5678)$ (see Example 2.1.14). Let $\chi : \mathbb{C} \to \mathbb{C}$ be complex conjugation (see Example 5.3.4). Then χ restricts to a permutation of S, hence defines an automorphism of $\mathbb{Q}(\zeta\alpha)$. Based on the ordering of S in (4.1), $\chi = (17)(28)(35)(46)$ is the disjoint cycle representation of χ. By direct computation, we see that $\tau\chi = (1836)(2547) = \chi\tau$. By Exercise 2.5.19, τ and χ generate an abelian group, call it G, isomorphic to $\mathbb{Z}/4 \oplus \mathbb{Z}/2$. Since $\dim_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha)) = 8 = [G : 1]$, by Proposition 5.3.12, $\mathbb{Q}(\zeta\alpha)$ is Galois over $\mathbb{Q}$ and the Galois group is $G = \langle\tau, \chi\rangle$. This also shows $G = \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha))$.

EXAMPLE 5.4.8. This is a generalization of Example 5.4.7. In this example we construct a Galois extension over $\mathbb{Q}$ such that the Galois group is isomorphic to the group of units in $\mathbb{Z}/(2^{n+1})$. As in Example 2.1.3, the set of invertible elements in the ring $\mathbb{Z}/(2^{n+1})$ is denoted $U_{2^{n+1}}$ and the order of this group is 2^n. Let a be a positive odd integer and $n \ge 2$. Let $f = x^{2^n} + a^{2^{n-1}}$. When $n = 3$, this example agrees with Example 5.4.7. By Exercise 3.7.12, f is irreducible over $\mathbb{Q}$. Let ζ be the complex number $e^{2\pi i/2^{n+1}}$, a primitive 2^{n+1}th root of unity. Then $\zeta^{2^{n+1}} = 1$ and $\zeta^{2^n} = -1$. Let α be the positive real number such that $\alpha^2 = a$. For any integer k,

$$f(\zeta^{2k-1}\alpha) = (\zeta^{2k-1}\alpha)^{2^n} + a^{2^{n-1}} = \zeta^{-2^n}(\zeta^{2^{n+1}})^k\alpha^{2^n} + a^{2^{n-1}} = -a^{2^{n-1}} + a^{2^{n-1}} = 0.$$

Therefore the 2^n roots of f in $\mathbb{C}$ are

$$S = \{\zeta^{2k-1}\alpha \mid 1 \le k \le 2^n\} = \{\zeta\alpha, \zeta^3\alpha, \ldots, \zeta^{2^{n+1}-1}\alpha\}.$$

By Theorem 5.1.4, the set

$$\{(\zeta\alpha)^j \mid 0 \le j < 2^n\} = \{1, \zeta\alpha, (\zeta\alpha)^2, \ldots, (\zeta\alpha)^{2^n-1}\}$$

is a basis for $\mathbb{Q}(\zeta\alpha)$ as a $\mathbb{Q}$-vector space. Since $(\zeta\alpha)^{2k+1} = \zeta^{2k+1}a^k\alpha$, we see that $S \subseteq \mathbb{Q}(\zeta\alpha)$. Hence $\mathbb{Q}(\zeta\alpha)$ is a splitting field for f. Let t be an arbitrary odd integer. By Corollary 5.1.7 applied to $\zeta\alpha$ and $\zeta^t\alpha$, there is an automorphism $\tau_t \in \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha))$ such that $\tau_t(\zeta\alpha) = \zeta^t\alpha$. Let s be another odd integer. Since ζ is a primitive 2^{n+1}th root of unity, Proposition 5.3.2 (2) implies that $\tau_t = \tau_s$ if and only if $s \equiv t \pmod{2^{n+1}}$. Since $\zeta^2\alpha^2 = \zeta^2 a$, it follows that $\zeta^2 \in \mathbb{Q}(\zeta\alpha)$. We have

$$\tau_t(\zeta^2) = \tau_t((\zeta\alpha)^2a^{-1}) = \tau_t(\zeta\alpha)^2a^{-1} = (\zeta^t\alpha)^2a^{-1} = (\zeta^{2t}a)a^{-1} = \zeta^{2t}.$$

Using this, we see that

$$\tau_t(\zeta^{2k+1}\alpha) = \tau_t((\zeta^2)^k\zeta\alpha) = (\zeta^{2t})^k(\zeta^t\alpha) = (\zeta^{2k+1})^t\alpha$$

and

$$\tau_s\tau_t(\zeta\alpha) = \tau_s(\zeta^t\alpha) = \zeta^{ts}\alpha = \tau_{ts}(\zeta\alpha).$$

Let σ denote an arbitrary automorphism in $\mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha))$. Then Proposition 5.3.2 (1) implies $\sigma(\zeta\alpha) = \zeta^t\alpha$ for a unique $t \in \{1,3,\ldots,2^{n+1}-1\}$. By Proposition 5.3.2 (2), σ is equal to τ_t. The computations above show that the assignment $\theta(t) = \tau_t$ defines an isomorphism of groups $\theta : U_{2^{n+1}} \to \mathrm{Aut}_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha))$. Since $\dim_{\mathbb{Q}}(\mathbb{Q}(\zeta\alpha)) = 2^n$, Proposition 5.3.12 implies $\mathbb{Q}(\zeta\alpha)$ is Galois over $\mathbb{Q}$ and the Galois group is isomorphic to $U_{2^{n+1}}$. See [**4**, Theorem 5.8.4] for a related result concerning cyclotomic extensions.

The next proposition shows that for a Galois extension F/k, if f is an irreducible separable polynomial in $k[x]$, then the irreducible factors of f in $F[x]$ all have the same degree.

PROPOSITION 5.4.9. *Let F/k be a Galois extension of fields and f an irreducible separable polynomial in $k[x]$. If the unique factorization of f in $F[x]$ is $f = f_1 \cdots f_m$, then* $\deg f_1 = \deg f_2 = \cdots = \deg f_m$.

PROOF. We prove this in two steps.

Step 1: Suppose K/k is a Galois extension of fields with group G. Assume f splits in $K[x]$. Let N be a normal subgroup of G and assume $F = K^N$. We prove that the irreducible factors of f in $F[x]$ all have the same degree. Let $X = \{\alpha_1,\ldots,\alpha_n\}$ be the roots of f in K. If $L = k(X)$ is the splitting field for f in K, then L/k is Galois by Theorem 5.3.15. By Exercise 5.3.29, $\mathrm{Aut}_k(L)$ acts transitively on X. By Theorem 5.3.18, $\mathrm{Aut}_k(L)$ is a homomorphic image of G, hence G acts transitively on X. Let a,b be two arbitrary elements of X. Let $\tau \in G$ such that $\tau(a) = b$. Since N is normal, $\tau N = N\tau$. Therefore $\tau N a = N\tau a = Nb$. This shows the orbit containing a is in one-to-one correspondence with the orbit containing b. Let $O_1,\ldots,O_m$ be the orbits of N acting on X. Then $|O_1| = \cdots = |O_m|$. For each $1 \le i \le m$, set $f_i = \prod_{a\in O_i}(x-a)$. We have

$$\begin{aligned} f &= \prod_{a\in X}(x-a) \\ &= \prod_{i=1}^{m}\prod_{a\in O_i}(x-a) \\ &= f_1\cdots f_m. \end{aligned}$$

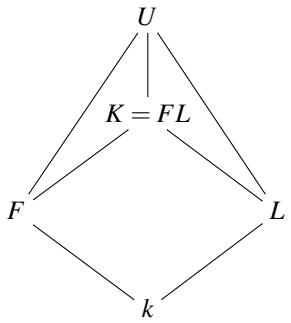
Since $\deg f_i = |O_i|$, all of the f_i have the same degree. Now we prove that each f_i is in $F[x]$. If $\tau \in N$, then $\tau O_i = O_i$, hence

$$\tau(f_i) = \prod_{a\in O_i}(x-\tau(a)) = \prod_{a\in O_i}(x-a) = f_i$$

so the coefficients of f_i are fixed by N. Hence $f_i \in F[x]$. Now we prove that each f_i is irreducible in $F[x]$. Fix one element of O_i, say a_i. If $p_i = \mathrm{Irr.poly}_F(a_i)$, then by Theorem 5.1.4 we have $p_i \mid f_i$. For each $\tau \in N$, $p_i(\tau a_i) = \tau(p_i(a_i)) = 0$ shows that every element of O_i is a root of p_i. Therefore, $\deg p_i \ge \deg f_i$. This proves $f_i = p_i$ and in particular, f_i is irreducible over F. We have proved that $f = f_1\cdots f_m$ is the factorization of f into irreducibles in the ring $F[x]$ and all of the factors f_i have the same degree.

Step 2. In the context of the proposition, assume F/k is a Galois extension. Let U/F be a splitting field for f over F. Let $X = \{\alpha_1,\ldots,\alpha_n\}$ be the roots of f in U. Let $L = k(X)$

be the splitting field for f over k in U. Then L/k is Galois by Theorem 5.3.15.



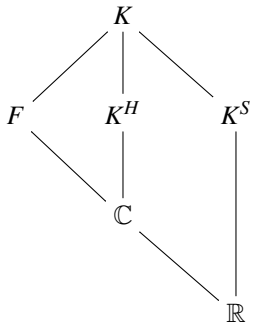
By Theorem 5.4.6, $K = FL$ is a Galois extension of k containing K. By Theorem 5.3.18, Step 2 reduces to Step 1. □

4.4. The Fundamental Theorem of Algebra. The purpose of this section is to apply Galois Theory and some facts about the completion of the metric space $\mathbb{R}$ to prove the Fundamental Theorem of Algebra.

As in Section 1.4, the field of real numbers is denoted $\mathbb{R}$ and the field of complex numbers is denoted $\mathbb{C}$. The proof of the Fundamental Theorem of Algebra utilizes results from Calculus. By Theorem 1.4.2, an irreducible polynomial of odd degree in $\mathbb{R}[x]$ is linear. By Proposition 1.4.3 (5), the ring $\mathbb{C}[x]$ contains no irreducible quadratic polynomial.

THEOREM 5.4.10. *The field of complex numbers is algebraically closed. In particular, an irreducible polynomial over $\mathbb{C}$ is linear.*

PROOF. By Lemma 5.2.1, we show that every irreducible polynomial over $\mathbb{C}$ is linear. Let F be a finite dimensional extension field of $\mathbb{C}$. By Theorem 5.2.4, it suffices to show that $F = \mathbb{C}$. Since F is a finite dimensional separable extension field of $\mathbb{R}$, by Corollary 5.3.17, there is a finite dimensional Galois extension $K/\mathbb{R}$ which contains F as an intermediate field. Let G be the Galois group of K over $\mathbb{R}$. Let S be a Sylow-2 subgroup of G. Then K^S is an extension field of $\mathbb{R}$ and $\dim_{\mathbb{R}} K^S$ is odd. If $\alpha \in K^S$, then $\dim_{\mathbb{R}} \mathbb{R}(\alpha)$ divides $\dim_{\mathbb{R}} K^S$, hence is odd. By Theorem 5.1.4, the degree of $\mathrm{Irr.poly}_{\mathbb{R}}(\alpha)$ is odd. By Theorem 1.4.2, an irreducible polynomial of odd degree in $\mathbb{R}[x]$ is linear. Therefore, $K^S = \mathbb{R}$. This proves $S = G$ is a 2-group. For sake of contradiction, assume $\mathrm{Aut}_{\mathbb{C}}(K)$ is a nontrivial 2-group. By Theorem 2.7.1, there exists a normal subgroup H of $\mathrm{Aut}_{\mathbb{C}}(K)$ of index 2. Then K^H is a field extension of $\mathbb{C}$ of degree 2. This is a contradiction, because by Proposition 1.4.3 (5), the ring $\mathbb{C}[x]$ contains no irreducible quadratic polynomial.



□

4.5. Exercises.

EXERCISE 5.4.11. Prove that (2) is equivalent to (3) in Theorem 5.4.5.

EXERCISE 5.4.12. Prove Theorem 5.4.6 (2).

5. The Trace Map and Norm Map

We show that to a Galois extension F/k are associated the trace map $T_k^F : F \to k$ and the norm map $N_k^F : F \to k$. By the left regular representation, F is a subring of $\mathrm{Hom}_k(F,F)$. Therefore we can make the group of linear functionals $\mathrm{Hom}_k(F,k)$ into an F-vector space. Since F/k is Galois, $\mathrm{Hom}_k(F,k)$ has dimension one over the field F and the trace map is a generator. We give a proof of Hilbert's Theorem 90 for the special case where $\mathrm{Aut}_k(F)$ is a cyclic group. Results from this short section will be applied in Section 5.6.

Let F/k be a Galois extension with finite group G. For $x \in F$, define

$$T_k^F(x) = \sum_{\sigma \in G} \sigma(x) \tag{5.1}$$

and

$$N_k^F(x) = \prod_{\sigma \in G} \sigma(x). \tag{5.2}$$

Since G is a group, for any $\tau \in G$,

$$\begin{aligned} \tau\Big(\sum_{\sigma\in G} \sigma(x)\Big) &= \sum_{\sigma\in G} \tau\sigma(x) \\ &= \sum_{\sigma\in G} \sigma(x) \end{aligned}$$

so the right hand side of (5.1) is fixed by every $\tau \in G$. Likewise,

$$\begin{aligned} \tau\Big(\prod_{\sigma\in G} \sigma(x)\Big) &= \prod_{\sigma\in G} \tau\sigma(x) \\ &= \prod_{\sigma\in G} \sigma(x) \end{aligned}$$

so the right hand side of (5.2) is fixed by G as well. Since $F^G = k$, this means that both T_k^F and N_k^F are mappings from F to k. We call the mapping T_k^F the *trace from F to k* and the mapping N_k^F is called the *norm from F to k*. If $x, y \in F$ and $a, b \in k$, then

$$\begin{aligned} T_k^F(ax+by) &= \sum_{\sigma\in G} \sigma(ax+by) \\ &= a\sum_{\sigma\in G} \sigma(x) + b\sum_{\sigma\in G} \sigma(y) \\ &= aT_k^F(x) + bT_k^F(y). \end{aligned}$$

Therefore, the trace is k-linear and represents an element of $\mathrm{Hom}_k(F,k)$. Also

$$\begin{aligned} N_k^F(xy) &= \prod_{\sigma\in G} \sigma(xy) \\ &= \prod_{\sigma\in G} \sigma(x) \prod_{\sigma\in G} \sigma(y) \\ &= N_k^F(x) N_k^F(y). \end{aligned}$$

Hence, the norm induces a homomorphism of multiplicative groups $F^* \to k^*$.

LEMMA 5.5.1. *Let F/k be a Galois extension of fields with finite group G. Let $\mathrm{Hom}_k(F,F)$ be the ring of k-linear transformations of F as a k-vector space.*

(1) There is a one-to-one homomorphism of rings $\lambda : F \to \mathrm{Hom}_k(F,F)$ defined by $\lambda(a) = \ell_a$, where ℓ_a is "left multiplication by a". That is, $\ell_a(x) = ax$. The homomorphism λ is called the left regular representation of F in $\mathrm{Hom}_k(F,F)$.

(2) If $n = [G:1]$, then $\mathrm{Hom}_k(F,F)$ is an F-vector space of dimension n and $\{\sigma \mid \sigma \in G\}$ is a basis.

(3) There exists $c \in F$ such that $T_k^F(c) = 1$.

(4) $\mathrm{Hom}_k(F,k)$ is an F-vector space of dimension 1 and $\{T_k^F\}$ is a basis.

PROOF. (1) and (2): The field F is a k-algebra, hence it acts as a ring of k-linear transformations on itself. That is, if $a \in F$, then $\ell_a(x) = ax$ defines a k-linear transformation: $\ell_a : F \to F$. The reader should verify that $a \mapsto \ell_a$ defines a homomorphism of rings $\lambda : F \to \mathrm{Hom}_k(F,F)$. Since F is a field, the image of λ is a commutative subring of $\mathrm{Hom}_k(F,F)$. By Example 4.1.4 (3), multiplication in $\mathrm{Hom}_k(F,F)$ makes $\mathrm{Hom}_k(F,F)$ into an F-vector space. By Theorem 5.3.5, G is a basis for $\mathrm{Hom}_k(F,F)$ as an F-vector space.

(3): By Theorem 5.3.5, there exists $y \in F$ such that $x = \sum_{\sigma\in G} \sigma(y) \neq 0$. Since G is a group, $\tau(x) = x$ for every $\tau \in G$. Therefore, $x \in F^G = k$. Define $c = x^{-1}y$. Then $T_k^F(c) = \sum_{\sigma\in G} \sigma(x^{-1}y) = x^{-1}\sum_{\sigma\in G} \sigma(y) = 1$.

(4): Using λ we can turn $\mathrm{Hom}_k(F,k)$ into an F-vector space. For every $f \in \mathrm{Hom}_k(F,k)$ and $a \in F$, define af to be $f \circ \ell_a$. By Proposition 4.2.8, $\mathrm{Hom}_k(F,k)$ is an F-vector space of dimension one. As an F-vector space, any nonzero element $f \in \mathrm{Hom}_k(F,k)$ is a generator. By (3), the trace mapping T_k^F is a generator for $\mathrm{Hom}_k(F,k)$ over F. This implies for every $f \in \mathrm{Hom}_k(F,k)$ there is a unique $\alpha \in F$ such that $f(x) = T_k^F(\alpha x)$ for all $x \in F$. The mapping $F \to \mathrm{Hom}_k(F,k)$ given by $\alpha \mapsto T_k^F \circ \ell_\alpha$ is an isomorphism of k-vector spaces. □

PROPOSITION 5.5.2. *Suppose F/k is G-Galois where the group G has order $[G:1] = n$. Then there exist elements $a_1,\ldots,a_n,y_1,\ldots,y_n$ in F such that*

(1) $T_k^F(y_j a_i) = \delta_{ij}$ (Kronecker delta), and

(2) for each $\sigma \in G$: $a_1\sigma(y_1) + \cdots + a_n\sigma(y_n) = \begin{cases} 1 & \text{if } \sigma = 1 \\ 0 & \text{if } \sigma \neq 1 \end{cases}$.

PROOF. Let $\{a_1,\ldots,a_n\}$ be a k-basis for F. For each $j = 1,2,\ldots,n$, let $f_j : F \to k$ be the projection onto coordinate j. That is, $f_j(a_i) = \delta_{ij}$ (Kronecker delta). For each $x \in F$,

$$x = \sum_{j=1}^{n} f_j(x)a_j.$$

We say $\{(a_j,f_j) \mid j = 1,\ldots,n\}$ is a *dual basis* for F. By Lemma 5.5.1, T_k^F is a generator for $\mathrm{Hom}_k(F,k)$ over F. There exist unique $y_1,\ldots,y_n$ in F such that for each $x \in F$, $f_j(x) = T_k^F(y_j x) = \sum_{\sigma\in G} \sigma(y_j x)$. Part (1) follows by substituting $x = a_i$. Combining these facts,

$$\begin{aligned} x &= \sum_{j=1}^{n} f_j(x)a_j \\ &= \sum_{j=1}^{n}\sum_{\sigma\in G} \sigma(y_j x)a_j \\ &= \sum_{\sigma\in G}\Big(\sigma(x)\sum_{j=1}^{n}\sigma(y_j)a_j\Big). \end{aligned}$$

By Lemma 5.5.1, G is a basis for $\mathrm{Hom}_k(F,F)$ over F. Therefore, $\sum_{j=1}^n \sigma(y_j)a_j = \delta_{\sigma,1}$, which is (2). □

LEMMA 5.5.3. *Suppose F/k is a Galois extension of fields with finite group G. If H is a normal subgroup of G and $E = F^H$, then $T_k^F = T_k^E \circ T_E^F$ and $N_k^F = N_k^E \circ N_E^F$.*

PROOF. Let $x \in F$. Then

$$\begin{aligned} T_k^E\Big(T_E^F(x)\Big) &= T_k^E\Big(\sum_{\sigma\in H}\sigma(x)\Big) \\ &= \sum_{\tau\in G/H}\tau\Big(\sum_{\sigma\in H}\sigma(x)\Big) \\ &= \sum_{\tau\in G/H}\sum_{\sigma\in H}\tau\sigma(x) \\ &= \sum_{\rho\in G}\rho(x) \\ &= T_k^F(x). \end{aligned}$$

The proof of the second identity is left to the reader. □

For generalizations of Theorem 5.5.4, see [**4**, Theorem 11.5.25]

THEOREM 5.5.4. *(Hilbert's Theorem 90) Let F/k be a Galois extension of fields with finite group G. Assume $G = \langle\sigma\rangle$ is cyclic and $u \in F$. Then*

(1) $T_k^F(u) = 0$ if and only if $u = v - \sigma(v)$ for some $v \in F$.
(2) $N_k^F(u) = 1$ if and only if $u = v/\sigma(v)$ for some $v \in F^*$.

PROOF. Throughout the proof, assume $G = \{1,\sigma,\sigma^2,\ldots,\sigma^{n-1}\}$ and $\sigma^n = 1$.

(1): If $v \in F$, then $T(\sigma(v)) = \sum_{\tau\in G}\tau\sigma(v) = \sum_{\rho\in G}\rho(v) = T(v)$. It follows that $T(v - \sigma(v)) = 0$. Conversely, assume $T(u) = 0$. By Lemma 5.5.1 (3), there exists $w \in F$ with $T(w) = 1$. Starting with

$$\begin{aligned} v = uw + (u+\sigma(u))\sigma(w) + (u+\sigma(u)+\sigma^2(u))\sigma^2(w) + \ldots \\ + (u+\sigma(u)+\sigma^2(u)+\cdots+\sigma^{n-2}(u))\sigma^{n-2}(w), \end{aligned}$$

apply σ to get

$$\begin{aligned} \sigma(v) = \sigma(u)\sigma(w) + (\sigma(u)+\sigma^2(u))\sigma^2(w) + \ldots \\ + (\sigma(u)+\sigma^2(u)+\cdots+\sigma^{n-1}(u))\sigma^{n-1}(w). \end{aligned}$$

Subtract $\sigma(v)$ from v. Use the identities $T(u) = u+\sigma(u)+\cdots+\sigma^{n-1}(u) = 0$ and $T(w) = 1$ to simplify

$$\begin{aligned} v - \sigma(v) &= uw + u\sigma(w) + u\sigma^2(w) + \cdots + u\sigma^{n-2}(w) \\ &\quad - \big(\sigma(u)+\sigma^2(u)+\cdots+\sigma^{n-1}(u)\big)\,\sigma^{n-1}(w) \\ &= u\big((w+\sigma(w)+\sigma^2(w)+\cdots+\sigma^{n-2}(w)\big) - (-u)\sigma^{n-1}(w) \\ &= u\big((w+\sigma(w)+\sigma^2(w)+\cdots+\sigma^{n-2}(w)+\sigma^{n-1}(w)\big) \\ &= uT(w) = u. \end{aligned}$$

(2): If $v \in F^*$, then $N(\sigma(v)) = \prod_{\tau\in G}\tau\sigma(v) = N(v)$. This shows $N(v/\sigma(v)) = 1$. Conversely, assume $N(u) = 1$. By Theorem 5.3.5 we know that

$$v = ux + u\sigma(u)\sigma(x) + u\sigma(u)\sigma^2(u)\sigma^2(x) + \cdots + u\sigma(u)\sigma^2(u)\cdots\sigma^{n-1}(u)\sigma^{n-1}(x)$$

is nonzero for some $x \in F$. In this case, we have

$$u^{-1}v = x + \sigma(u)\sigma(x) + \sigma(u)\sigma^2(u)\sigma^2(x) + \cdots + \sigma(u)\sigma^2(u)\cdots\sigma^{n-1}(u)\sigma^{n-1}(x)$$

and

$$\begin{aligned}\sigma(v) &= \sigma(u)\sigma(x) + \sigma(u)\sigma^2(u)\sigma^2(x) + \cdots + \sigma(u)\sigma^2(u)\cdots\sigma^n(u)\sigma^n(x)\\ &= \sigma(u)\sigma(x) + \sigma(u)\sigma^2(u)\sigma^2(x) + \cdots + N(u)x\\ &= \sigma(u)\sigma(x) + \sigma(u)\sigma^2(u)\sigma^2(x) + \cdots + x.\end{aligned}$$

This shows $\sigma(v) = u^{-1}v$, hence $u = v/\sigma(v)$. □

5.1. Exercises.

EXERCISE 5.5.5. Let k be a field. Show that for any $n \geq 1$ there exists a polynomial $f \in F[x]$ of degree n such that f has no repeated roots.

EXERCISE 5.5.6. Let F/k be a Galois extension of fields with finite group G. Assume $G = \langle\sigma\rangle$ is cyclic.

(1) Show that the function $D : F^* \to F^*$ defined by $D(u) = u/\sigma(u)$ is a homomorphism of abelian groups.
(2) Show that the kernel of D is k^*, and the image of D is the kernel of $N_k^F : F^* \to k^*$.

EXERCISE 5.5.7. For the cyclic Galois extension $\mathbb{C}/\mathbb{R}$ of degree two, determine the image of the norm map $N_{\mathbb{R}}^{\mathbb{C}} : \mathbb{C}^* \to \mathbb{R}^*$ and show that it is a subgroup of $\mathbb{R}^*$ of index two.

EXERCISE 5.5.8. Let F/k be a Galois extension of fields with finite group G. Assume $G = \langle\sigma\rangle$ is cyclic.

(1) Show that the function $D : F \to F$ defined by $D(x) = x - \sigma(x)$ is a homomorphism of additive abelian groups.
(2) Show that the kernel of D is k, and the image of D is the kernel of the trace map $T_k^F : F \to k$.

EXERCISE 5.5.9. Let F/k be an extension of fields and assume $\dim_k F = n$ is finite. As in Lemma 5.5.1, the left regular representation $\lambda : F \to \mathrm{Hom}_k(F,F)$ makes $\mathrm{Hom}_k(F,F)$ into a left F-vector space. Prove:

(1) $\dim_F(\mathrm{Hom}_k(F,F)) = n$.
(2) If $\{v_1,\ldots,v_n\}$ is a k-basis for F and $\{\phi_1,\ldots,\phi_n\}$ is an F-basis for $\mathrm{Hom}_k(F,F)$, then the matrix $(\phi_i(v_j))$ is invertible in $M_n(F)$.
(3) If F/k is a Galois extension of fields with group $G = \{\sigma_1,\ldots,\sigma_n\}$, then the matrix $(\sigma_i(v_j))$ in $M_n(F)$ is invertible.

6. Cyclic Galois Extensions

We say a finite Galois extension of fields F/k is *cyclic of degree n* if the group $\mathrm{Aut}_k(F)$ is a cyclic group of order n.

THEOREM 5.6.1. *(The Normal Basis Theorem) Let F/k be a cyclic Galois extension of degree n with group $\mathrm{Aut}_k(F) = \langle\sigma\rangle$. Then there exists $\alpha \in F$ such that the set $\{\alpha, \sigma(\alpha), \sigma^2(\alpha), \ldots, \sigma^{n-1}(\alpha)\}$ is a basis for F as a k-vector space. We call the basis $\{\alpha, \sigma(\alpha), \sigma^2(\alpha), \ldots, \sigma^{n-1}(\alpha)\}$ a* normal basis for F/k.

PROOF. We have $\dim_k(F) = n$. View $1, \sigma, \sigma^2, \ldots, \sigma^{n-1}$ as elements of $\mathrm{Hom}_k(F,F)$. Then $\mathrm{char.poly}_k(\sigma)$ has degree n (see Definition 4.6.13). Since $\mathrm{Aut}_k(F) = \langle\sigma\rangle$ has order n, the minimal polynomial of σ divides $x^n - 1$. By Theorem 5.3.5, the automorphisms $1, \sigma, \sigma^2, \ldots, \sigma^{n-1}$ are linearly independent over k, so the degree of $\mathrm{min.poly}_k(\sigma)$ is at least n. Therefore, $\mathrm{min.poly}_k(\sigma) = x^n - 1$. Since the minimal polynomial and the characteristic polynomial of σ both have degree n, this implies they are equal. By Theorem 4.6.14, F is a cyclic $k[\sigma]$-module. By Theorem 4.5.31, there exists $\alpha \in F$ such that the set $\{\alpha, \sigma(\alpha), \sigma^2(\alpha), \ldots, \sigma^{n-1}(\alpha)\}$ is a k-basis for F. □

6.1. Finite Fields. A finite field has positive characteristic and is finite dimensional over its prime subfield. We prove in Theorem 5.6.4 (9) that a finite extension of finite fields is a cyclic extension.

LEMMA 5.6.2. *Let F be a field and assume* $\mathrm{char}\, F = p$ *is positive. For any $r > 0$, the mapping $\varphi : F \to F$ defined by $x \mapsto x^{p^r}$ is a homomorphism of fields. If F is finite, then φ is an automorphism of F. If $r = 1$, then φ is called the* Frobenius homomorphism.

PROOF. This follows from Exercise 3.2.31. □

LEMMA 5.6.3. *For each prime number p and for every $n \geq 1$, there exists a field F of order p^n.*

PROOF. Let k denote the field $\mathbb{Z}/p$. Let $f = x^{p^n} - x \in k[x]$. Let F be the splitting field of f over k. Since $f' = -1$, by Theorem 3.6.15, f has no multiple roots in F. Therefore, f is separable and there are p^n distinct roots of f in F. Let $\varphi : F \to F$ be the automorphism of F defined by $x \mapsto x^{p^n}$. If $u \in F$ is a root of f, then $\varphi(u) = u$. By Exercise 5.3.21, the prime field k is fixed by φ. Since F is generated over k by roots of f, F is fixed point-wise by φ. Every u in F is a root of f, and F has order p^n. □

THEOREM 5.6.4. *(Fundamental Theorem on Finite Fields) Let F be a finite field with* $\mathrm{char}\, F = p$. *Let k be the prime subfield of F and $n = \dim_k(F)$.*

(1) The group of units of F is a cyclic group.
(2) $F = k(u)$ is a simple extension, for some $u \in F$.
(3) The order of F is p^n.
(4) F is the splitting field for the separable polynomial $x^{p^n} - x$ over k.
(5) F/k is a separable extension.
(6) Any two finite fields of order p^n are isomorphic as fields.
(7) F/k is a Galois extension.
(8) The Galois group $\mathrm{Aut}_k(F)$ is cyclic of order n and is generated by the Frobenius homomorphism $\varphi : F \to F$ defined by $\varphi(x) = x^p$.
(9) If d is a positive divisor of n, then $E = \{u \in F \mid u^{p^d} = u\}$ is an intermediate field of F/k which satisfies the following.
  (a) $\dim_E(F) = n/d$, and $\dim_k(E) = d$.
  (b) If φ is the generator for $\mathrm{Aut}_k(F)$, then $\mathrm{Aut}_E(F)$ is the cyclic subgroup generated by φ^d.
  (c) E/k is Galois and $\mathrm{Aut}_k(E)$ is the cyclic group of order d generated by the restriction $\varphi|_E$.
(10) If E is an intermediate field of F/k, and $d = \dim_k(E)$, then d divides n and E is the field described in Part (9).

PROOF. Parts (1) – (6) are from Theorem 5.2.12. Parts (7) and (8) are from Example 5.3.13. The proofs of Parts (9) and (10) follow straight from Theorem 5.3.18 and Part (8). □

6.1.1. *Irreducible Polynomials over Finite Fields.* Throughout this section, p will be a fixed prime number and $\mathbb{F}_p = \mathbb{Z}/p$ is the prime field of order p.

THEOREM 5.6.5. *The factorization of the polynomial $x^{p^n} - x$ in $\mathbb{F}_p[x]$ into irreducible factors is equal to the product of all the monic irreducible polynomials of degree d where d runs through all divisors of n.*

PROOF. Is left to the reader. □

THEOREM 5.6.6. *Let $\psi(n)$ denote the number of distinct monic irreducible polynomials of degree n in $\mathbb{F}_p$.*

(1) If μ is the Möbius function, then $\psi(n) = \dfrac{1}{n}\sum_{d|n}\mu(d)p^{n/d} = \dfrac{1}{n}\sum_{d|n}\mu\left(\dfrac{n}{d}\right)p^d$.

(2) $\psi(n) > \dfrac{p^n}{2n}$.

PROOF. (1): By Theorem 5.6.5, $p^n = \sum_{d|n} d\psi(d)$. Now apply the Möbius Inversion Formula (Theorem 1.2.16).

(2): The reader should verify the identities:

$$\begin{aligned} n\psi(n) &= p^n + \sum_{d|n,d<n} \mu\left(\frac{n}{d}\right)p^d \\ &\geq p^n - \sum_{d|n,d<n} p^d \\ &\geq p^n - \sum_{1\leq d\leq n/2} p^d \\ &\geq p^n - p^{\lfloor n/2\rfloor+1} \end{aligned}$$

where $\lfloor n/2\rfloor$ is the greatest integer less than $n/2$. If $n > 2$, then $\lfloor n/2\rfloor + 1 \leq n-1$, so

$$\psi(n) > \frac{1}{n}\left(p^n - p^{n-1}\right) = \frac{p^n}{n}\left(1 - \frac{1}{p}\right) \geq \frac{p^n}{2n}.$$

If $n = 2$, the formula can be derived from $\psi(2) = (1/2)(p^2 - p)$. □

6.2. Exercises.

EXERCISE 5.6.7. Prove Theorem 5.6.5.

EXERCISE 5.6.8. Let K be a finite field of order p^d. As in Theorem 5.6.6, let $\psi(n)$ be the number of irreducible monic polynomials of degree n in $\mathbb{F}_p[x]$. If $d \mid n$, show that there are at least $\psi(n)$ irreducible monic polynomials of degree n/d in $K[x]$.

EXERCISE 5.6.9. Let k be a finite field and K/k a finite dimensional extension of fields, with $\dim_k K = d$. Let n be an arbitrary positive integer and $A = K \oplus \cdots \oplus K$ the direct sum of n copies of K.

(1) Show that if there exists a surjective k-algebra homomorphism $f : k[x] \to A$, then there exist at least n distinct irreducible monic polynomials in $k[x]$ of degree d.

(2) Find an example of k and A such that the k-algebra A is not the homomorphic image of $k[x]$.
(3) Show that for some integer $m \geq 1$, there exist n distinct irreducible monic polynomials $h_1, \ldots, h_n$ in $k[x]$ such that each h_i has degree md.
(4) Show that for some integer $m \geq 1$, if F/k is a finite extension field with $\dim_k F = md$, then the direct sum $F \oplus \cdots \oplus F$ of n copies of F is the homomorphic image of $k[x]$. Show that m can be chosen to be relatively prime to d.
(5) Show that there is a separable polynomial $g \in k[x]$ such that A is isomorphic to a subalgebra of $k[x]/(g)$.

EXERCISE 5.6.10. Classify up to isomorphism all finite rings of order four. For a generalization of this result to rings of order p^2, p a prime number, see Exercise 5.6.11. The reader interested in rings that do not necessarily contain a unit element is referred to the classification obtained in [**12**].

EXERCISE 5.6.11. Let p be a prime number and A a finite ring of order p^2.

(1) Prove that either A is isomorphic to $\mathbb{Z}/(p^2)$, or the characteristic of A is p and A is isomorphic as $\mathbb{Z}/p$-algebras to $(\mathbb{Z}/p)[x]/(\phi)$, for some monic quadratic polynomial ϕ with coefficients in the field $\mathbb{Z}/p$.
(2) Prove that A is commutative.
(3) Prove that A is isomorphic to exactly one of the following four rings:
   (a) $\mathbb{Z}/(p^2)$ (if $\text{char}(A) = p^2$).
   (b) $\mathbb{Z}/p \oplus \mathbb{Z}/p$ (if $\text{char}(A) = p$ and ϕ factors and is separable).
   (c) $(\mathbb{Z}/p)[x]/(x^2)$ (if $\text{char}(A) = p$ and ϕ is a square).
   (d) a finite field of order p^2 (if $\text{char}(A) = p$ and ϕ is irreducible).

EXERCISE 5.6.12. If F/k is an extension of finite fields, show that the image of the norm map $N_k^F : F^* \to k^*$ is equal to k^*.

6.3. Artin-Schreier Theorem.

EXAMPLE 5.6.13. Let k be a field of positive characteristic p. For any $a \in k$, the polynomial $f = x^p - x - a \in k[x]$ is separable over k. To see this, assume u is a root of f in any extension field F/k. Let $i \in \mathbb{Z}/p$ be any element of the prime field of k. Then $f(u+i) = (u+i)^p - (u+i) - a = u^p + i - u - i - a = f(u) = 0$. Therefore, f has p distinct roots in F, namely $u, u+1, \ldots, u+p-1$.

THEOREM 5.6.14. *(Artin-Schreier) Suppose k is a field of positive characteristic p.*

(1) If F/k is a cyclic Galois extension of degree p, then there exists $a \in k$ such that $f = x^p - x - a$ is an irreducible separable polynomial over k and F is the splitting field for f over k. In this case $F = k(u)$, where u is any root of f.
(2) If $a \in k$ and $f = x^p - x - a$, then
   (a) f is separable, and
   (b) either f is irreducible over k, or splits in $k[x]$.
(3) If $a \in k$ and $f = x^p - x - a$ is irreducible over k, then
   (a) $F = k[x]/(f)$ is a splitting field for f, and
   (b) F/k is a cyclic Galois extension of k of degree p.

PROOF. (1): Let $G = \text{Aut}_k(F) = \langle \sigma \rangle$. Since G is simple and abelian (Exercise 2.2.28), there are no proper intermediate fields for F/k. Since $\text{char}(k) = \dim_k(F) = p$, $T_k^F(1) = p = 0$. By Theorem 5.5.4, there is $v \in F$ such that $v - \sigma(v) = 1$. If $u = -v$, then $\sigma(u) = 1 + u$. This shows $u \notin k$, hence $F = k(u)$. Note that $\sigma(u^p) = (\sigma(u))^p = (1+u)^p = 1 + u^p$, and

$\sigma(u^p-u)=\sigma(u^p)-\sigma(u)=(1+u^p)-(u+1)=u^p-u$. If $a=u^p-u$, then $a\in k$ and u satisfies the polynomial $f=x^p-x-a$. Since the dimension of $k(u)$ over k is p, this implies f is equal to the irreducible polynomial of u. By Example 5.6.13, f is separable and splits in F.

(2): Let $f=x^p-x-a$ in $k[x]$. Let F be a splitting field for f. As was shown in Example 5.6.13, f is separable and if $u\in F$ is a root of f, then the p distinct roots of f are $u,u+1,\ldots,u+p-1$, hence $F=k(u)$. By Theorem 5.3.15, F/k is a Galois extension. For any τ in $\mathrm{Aut}_k(F)$, by Proposition 5.3.2, $\tau(u)$ is a root of f. Thus, $\tau(u)-u$ is an element of the prime field $\mathbb{Z}/p$. Define a function $\theta:\mathrm{Aut}_k(F)\to\mathbb{Z}/p$ by $\theta(\tau)=\tau(u)-u$. If σ is another element of $\mathrm{Aut}_k(F)$, then $\sigma(\tau(u)-u)=\tau(u)-u$. Hence $\sigma\tau(u)-\sigma(u)=\tau(u)-u$. From this we see that

$$\sigma\tau(u)-u=\sigma(u)+\tau(u)-u-u. \tag{6.1}$$

The left hand side of (6.1) is $\theta(\sigma\tau)$, the right hand side is $\theta(\sigma)+\theta(\tau)$. This shows θ is a homomorphism from the group $\mathrm{Aut}_k(F)$ to the additive cyclic group $\mathbb{Z}/p$. By Proposition 5.3.2, θ is one-to-one. Since $\mathbb{Z}/p$ is a simple group, either $\mathrm{Aut}_k(F)$ has order 1 or p. By Theorem 5.3.18, if $\mathrm{Aut}_k(F)$ has order 1, then $F=k$ and f splits in $k[x]$. If $\mathrm{Aut}_k(F)$ has order p, then $\dim_k(F)=p$. Since $F=k(u)$, by Theorem 5.1.4, $\mathrm{Irr.poly}_k(u)$ has degree p. Therefore, $f=\mathrm{Irr.poly}_k(u)$.

(3): This follows from Part (2). □

6.4. Kummer Theory. If $\zeta\in k^*$ and ζ generates a subgroup of order n in k^*, then we say ζ is a *primitive nth root of* 1 *in* k and write $\zeta=\sqrt[n]{1}$. There are at most n solutions to x^n-1 in k, so the subgroup $\langle\zeta\rangle$ has $\varphi(n)$ generators. That is, if k contains a primitive nth root of 1, then k contains $\varphi(n)$ primitive nth roots of 1. A cyclic extension F/k of degree n is called a *Kummer extension* if $\sqrt[n]{1}\in k$.

THEOREM 5.6.15. *Let $n>0$ and assume k is a field containing a primitive nth root of 1. The following are equivalent.*

1. *F/k is a cyclic Galois extension of degree d, for some positive divisor d of n.*
2. *F is a splitting field over k of x^n-a for some $a\in k^*$.*
3. *F is a splitting field over k of x^d-a for some $a\in k^*$ and some positive divisor d of n.*

PROOF. Throughout the proof, let $\zeta=\sqrt[n]{1}$ be a primitive nth root of 1 in k.

(2) implies (1): Let α be a root of x^n-a in F. For each $i\geq 0$ we have $\left(\zeta^i\alpha\right)^n=\left(\zeta^n\right)^i\alpha^n=a$, so the roots of x^n-a in F are $\{\zeta^i\alpha\mid 0\leq i<n\}$. This shows x^n-a is separable. Also, since $\zeta\in k$, this implies $F=k(\alpha)$ is a simple extension. If $\sigma\in G=\mathrm{Aut}_k(F)$, then $\sigma(\alpha)=\zeta^i\alpha$ for some i such that $0\leq i<n$. As σ runs through the nonidentity elements of G, consider the positive numbers i such that $\sigma(\alpha)=\zeta^i\alpha$ and pick the smallest. Fix $\sigma\in G$, such that $\sigma(\alpha)=\zeta^i\alpha$ and i is minimal. We prove that G is generated by σ. Let τ be any element of G. Then $\tau(\alpha)=\zeta^j\alpha$ and we can assume $0<i\leq j<n$. Dividing, $j=iq+r$, where $0\leq r<i$. Now $\sigma^q(\alpha)=\zeta^{qi}\alpha$. Therefore, $\sigma^{-q}\tau(\alpha)=\sigma^{-q}(\zeta^j\alpha)=\zeta^j\sigma^{-q}(\alpha)=\zeta^j\zeta^{-qi}\alpha=\zeta^r\alpha$. By the choice of i we conclude that $r=0$, so $\tau=\sigma^q$. The order of G is equal to the order of ζ^i, which is a divisor of n.

(3) implies (2): Assume F is the splitting field of x^d-a where d is a divisor of n, and $a\in k$. Let $\rho=\zeta^{n/d}$. Then $\rho=\sqrt[d]{1}$. Let $\alpha\in F$ satisfy $\alpha^d=a$. Then x^d-a factors in $F[x]$ as $(x-\alpha)(x-\rho\alpha)\cdots(x-\rho^{d-1}\alpha)$. This implies $F=k(\alpha)$, because $\rho\in k$. Consider the

polynomial $x^n - a^{n/d}$. For any i such that $0 \leq i < n$ we see that $\left(\zeta^i \alpha\right)^n = \alpha^n = \left(\alpha^d\right)^{n/d} = a^{n/d}$. So $x^n - a^{n/d}$ splits in F.

(1) implies (3): Assume F/k is cyclic of degree d and that σ is a generator for $G = \mathrm{Aut}_k(F)$. Since $\rho = \zeta^{n/d} = \sqrt[d]{1}$ is in k, the norm of ρ is $N(\rho) = \rho^d = 1$. By Theorem 5.5.4, there is $u \in F^*$ such that $\rho = u/\sigma(u)$. Setting $v = u^{-1}$, we have $\rho = v^{-1}\sigma(v)$, or $\sigma(v) = \rho v$. Then $\sigma(v^d) = (\rho v)^d = v^d$. This says $v^d \in k$ and v satisfies the polynomial $x^d - v^d$. The roots of $x^d - v^d$ are $\{v, \rho v, \ldots, \rho^{d-1} v\}$. Note that $\sigma^i(v) = \rho^i v$, for all i such that $0 \leq i < d$. If f is the irreducible polynomial for v, then f has d roots in F. Therefore $\deg(f) = d$ and $f = x^d - v^d$. We have shown that F is the splitting field of f and $F = k(v)$. □

6.5. Cyclotomic Extensions. Let k be a field. We say F is a *cyclotomic extension* of k of order n if F is the splitting field over k of $x^n - 1$. If $\mathrm{char}\, k = p > 0$, then we can factor $n = p^e m$ where $(m, p) = 1$. Then $x^n - 1 = (x^m)^{p^e} - 1^{p^e} = (x^m - 1)^{p^e}$, so the splitting field of $x^n - 1$ is equal to the splitting field of $x^m - 1$. For this reason, we assume n is relatively prime to $\mathrm{char}\, k$ and $x^n - 1$ is separable. In the following, $\phi(n)$ denotes the Euler ϕ-function.

LEMMA 5.6.16. *Let k be any field. If m and n are positive integers and $m \mid n$, then $x^m - 1$ divides $x^n - 1$ in the ring $k[x]$. Conversely, if the characteristic of k does not divide m and $x^m - 1$ divides $x^n - 1$, then $m \mid n$.*

PROOF. Use Mathematical Induction on $n - m$. If $m = n$, then this is trivial. Assume $m < n$ and apply the Division Algorithm to write

$$x^n - 1 = (x^m - 1)x^{n-m} + (x^{n-m} - 1).$$

Since $m \mid n$ we have $m \mid (n - m)$. By Mathematical Induction, $x^m - 1$ divides $x^{n-m} - 1$. Therefore, $x^m - 1$ divides the right hand side.

For the converse, let F be a field extension of k containing all of the roots of $x^n - 1$. By hypothesis, we can factor $x^n - 1 = (x^m - 1)q(x)$ for some $q(x) \in k[x]$. If we let $f = x^m - 1$, then f splits over F. Since $\mathrm{char}\, k$ does not divide m, we have $\gcd(f, f') = 1$. By Theorem 3.6.15 (1), $f = x^m - 1$ has m distinct roots in F. By Corollary 3.6.9, the set of roots of $x^m - 1$ is a cyclic subgroup of F^* of order m. That is, there exists an element $\alpha \in F^*$ such that α has order m. Then $\alpha^n - 1 = (\alpha^m - 1)g(\alpha) = 0$ says $\alpha^n = 1$. By Lemma 2.2.16, we have $m \mid n$. □

THEOREM 5.6.17. *Let F be a cyclotomic extension of k of order n. If* $\mathrm{char}\, k = p > 1$, *assume $(n, p) = 1$. Then*

(1) $F = k(\zeta)$ where ζ is a primitive nth root of 1 *over k.*
(2) F is a Galois extension of k and $\mathrm{Aut}_k(F)$ is a subgroup of the group of units in $\mathbb{Z}/n$. The dimension $\dim_k(F)$ is a divisor of $\phi(n)$.

PROOF. (1): By assumption, $x^n - 1$ is separable, and the group μ_n of nth roots of unity in F is a cyclic group of order n, by Corollary 3.6.9. Let ζ be a primitive nth root of unity in F. Therefore $F = k(\zeta)$ is a simple extension.

(2): Since F is the splitting field of a separable polynomial, F/k is Galois by Theorem 5.3.15. The Galois group $G = \mathrm{Aut}_k(F)$ acts on the cyclic group of order n generated by ζ (Corollary 5.3.16). This defines a homomorphism $G \to \mathrm{Aut}(\langle\zeta\rangle)$. Since $F = k(\zeta)$, this mapping is one-to-one. By Theorem 2.3.27, the order of $\mathrm{Aut}(\langle\zeta\rangle)$ is $\phi(n)$. □

Let F be a cyclotomic extension of k of order n. If $\mathrm{char}\, k = p > 1$, assume $(n, p) = 1$. By Theorem 5.6.17 (1), the group μ_n of nth roots of unity in F is a cyclic group of order n.

There are $\phi(n)$ generators of μ_n. The *nth cyclotomic polynomial over k* is

$$\Phi_n(x) = (x-\zeta_1)\cdots(x-\zeta_{\phi(n)})$$

where $\zeta_1,\ldots,\zeta_{\phi(n)}$ are the $\phi(n)$ primitive nth roots of unity in μ_n. We have seen in Examples 5.2.9 and 3.7.8 that if p is a prime number and $k=\mathbb{Q}$, then $\Phi_p(x) = x^{p-1}+x^{p-2}+\cdots+x+1$ is irreducible in $\mathbb{Q}[x]$.

PROPOSITION 5.6.18. *Assume k is the prime subfield of F and F is a cyclotomic extension of k of order n. If* $\operatorname{char} k = p > 1$*, assume* $(n,p)=1$*. Then*

(1) $x^n-1 = \prod_{d|n}\Phi_d(x)$.
(2) $\Phi_n(x)\in k[x]$.
(3) If $k=\mathbb{Q}$*, then* $\Phi_n(x)\in\mathbb{Z}[x]$.

PROOF. (1): By Theorem 2.3.25, we can partition μ_n into disjoint subsets

$$\mu_n = \bigcup_{d|n}\{\zeta\in\mu_n \mid |\zeta| = d\}.$$

The set elements of order d in μ_n has cardinality $\phi(d)$. The corresponding factorization of x^n-1 is $x^n-1=\prod_{d|n}\Phi_d(x)$.

(2): The proof is by induction on n. For $n=1$, $\Phi_1(x)=x-1$ is in $k[x]$. Assume $n>1$ and that (2) is true for all $1\le m<n$. Define $g(x)=\prod_{\substack{d|n\\ d\neq n}}\Phi_d(x)$. By our induction hypothesis, $g(x)\in k[x]$. By (1), $x^n-1=g(x)\Phi_n(x)$. By the Division Algorithm, Theorem 3.6.3, $\Phi_n(x)\in k[x]$.

(3): In the proof of (2), by the induction hypothesis, $g(x)\in\mathbb{Z}[x]$. Moreover, $g(x)$ is monic, so Theorem 3.6.3 implies $\Phi_n(x)\in\mathbb{Z}[x]$. □

PROPOSITION 5.6.19. *If* $\Phi_n(x)$ *is the nth cyclotomic polynomial over* $\mathbb{Q}$*, then* $\Phi_n(x)$ *is irreducible.*

PROOF. Let F be a cyclotomic extension of order n over the field $\mathbb{Q}$ and $\Phi_n(x)$ the nth cyclotomic polynomial over $\mathbb{Q}$. We know from Proposition 5.6.18 that $\Phi_n(x)$ is a monic polynomial in $\mathbb{Z}[x]$ and has degree $\phi(n)$. By Theorem 3.7.4 (Gauss' Lemma) it suffices to show that $\Phi_n(x)$ is irreducible in $\mathbb{Z}[x]$. Let $f(x)$ be a monic irreducible factor of $\Phi_n(x)$ in $\mathbb{Z}[x]$ and write $\Phi_n(x)=f(x)g(x)$. To complete the proof, we show that $\Phi_n(x)=f(x)$. To do this, we show that $f(x)$ has degree $\phi(n)$. Let $\zeta\in F$ be a root of f. Then ζ is a root of $\Phi_n(x)$, hence is a primitive nth root of unity. By Theorem 2.3.25, a typical primitive nth root of unity is of the form ζ^d, where $0<d<n$ and $\gcd(d,n)=1$. We show that each such ζ^d is a root of f. We do this is several steps.

First let p be a prime divisor of d. Then ζ^p is a root of $\Phi_n(x)=f(x)g(x)$. We show ζ^p is a root of f. For contradiction's sake, assume $g(\zeta^p)=0$. Then ζ is a root of $g(x^p)$. Since $f(x)$ is irreducible, $f=\operatorname{Irr.poly}_{\mathbb{Q}}(\zeta)$ and by Theorem 5.1.4 we have $f(x)\mid g(x^p)$ in $\mathbb{Q}[x]$. By Theorem 3.6.3 applied over $\mathbb{Z}$ and $\mathbb{Q}$, we have $g(x^p)=f(x)h(x)$ where $h\in\mathbb{Z}[x]$. We apply Theorem 3.6.2 (1) to reduce the coefficients of the polynomials modulo p. The image of the polynomial $g(x^p)=f(x)h(x)$ under the natural map $\mathbb{Z}[x]\to\mathbb{Z}/(p)[x]$ will be denoted $[g(x^p)]=[f(x)][h(x)]$. The Frobenius homomorphism $\mathbb{Z}/(p)[x]\to\mathbb{Z}/(p)[x]$ of Exercise 3.2.31 fixes the field $\mathbb{Z}/(p)$, hence $[g(x^p)]=[g(x)]^p=[f(x)][h(x)]$. By unique factorization, some irreducible factor of $[f(x)]$ divides $[g(x)]$. By Proposition 5.6.18 (1), for some $q(x)\in\mathbb{Z}[x]$ we have $x^n-1=\Phi_n(x)q(x)=f(x)g(x)q(x)$. Reduce modulo p to get $x^n-1=[f(x)][g(x)][q(x)]$. Since $[f(x)]$ and $[g(x)]$ have a common factor, this proves x^n-1 is not separable, a contradiction. This proves ζ^p is a root of $f(x)$.

Now assume $0 < d < n$ and $\gcd(d,n) = 1$. We show that ζ^d is a root of f. Factor $d = p_1 \cdots p_m$ into a product of primes. If $m = 1$, then by the first step, ζ^d is a root of f. Inductively assume $m > 1$ and that $\zeta^{p_1 \cdots p_{m-1}}$ is a root of f. Then by the first step, $\zeta^d = (\zeta^{p_1 \cdots p_{m-1}})^{p_m}$ is a root of f. Since there are $\phi(n)$ choices for d, we have shown f has $\phi(n)$ roots, hence $\Phi_n(x) = f(x)$ is irreducible. □

COROLLARY 5.6.20. *Let F be a cyclotomic extension of order n over the field $\mathbb{Q}$ and $\Phi_n(x)$ the nth cyclotomic polynomial over $\mathbb{Q}$. Then the following are true.*

(1) If $\zeta \in F$ is a primitive nth root of unity, then $\Phi_n(x) = \text{Irr.poly}_{\mathbb{Q}}(\zeta)$.
(2) $F \cong \mathbb{Q}[x]/(\Phi_n)$.
(3) F is a Galois extension of $\mathbb{Q}$, the Galois group $\text{Aut}_{\mathbb{Q}}(F)$ is isomorphic to the group of units in the ring $\mathbb{Z}/(n)$, and $\dim_{\mathbb{Q}}(F) = \phi(n)$.

6.6. Radical Extensions. Throughout this section, all fields are tacitly assumed to have characteristic zero.

DEFINITION 5.6.21. Let k be a field. A *radical tower* over k is a tower of field extensions

$$k = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_n$$

and positive integers $r_1, \ldots, r_n$ such that $F_i = F_{i-1}(u_i)$ and $u_i^{r_i} \in F_{i-1}$. We say F_n is a *radical extension* of k. Notice that $F_i = k(u_1, \ldots, u_i)$, for $i = 1, \ldots, n$. If $f(x) \in k[x]$, we say f is *solvable by radicals* in case there is a radical extension F/k such that f splits over F.

LEMMA 5.6.22. *Let F/k be a finite dimensional separable extension of fields. Then there is a field K satisfying the following.*

(1) $k \subseteq F \subseteq K$ is a tower of field extensions.
(2) K/k is a Galois extension.
(3) There exist intermediate fields $F_1, \ldots, F_m$ of K/k such that
  (a) each F_i is isomorphic to F as a k-algebra, and
  (b) $K = F_1 F_2 \cdots F_m$.
(4) If F/k is a radical extension, then K/k is a radical extension.

PROOF. Write $F = k(u_1, \ldots, u_n)$. For each i in $\{1, \ldots, n\}$, let $f_i = \text{Irr.poly}_k(u_i)$. Let K be the splitting field for $f_1 \cdots f_n$ over F. By the Embedding Theorem, Corollary 5.3.17, the field K satisfies parts (1) and (2). We prove that K satisfies (3). Let $\alpha \in K$ be an arbitrary root of $f_1 \cdots f_n$. Then α is a root of f_i, for some i. By Theorem 5.1.5, there is a k-algebra isomorphism $\theta : k(u_i) \to k(\alpha)$. By Lemma 5.2.7, θ extends to a k-algebra automorphism $\bar{\theta} : K \to K$. Then $\bar{\theta}(F)$ is an intermediate field of K/k which is k-isomorphic to F and contains α. Since K/k is generated by the roots α of $f_1 \cdots f_n$, there is a finite number of fields of the form $\bar{\theta}(F)$ that generate K.

(4): We are given $F = k(u_1, \ldots, u_n)$, where $u_i^{r_i}$ is in $k(u_1, \ldots, u_{i-1})$. Let $F_1, \ldots, F_m$ be as in (3). For each i, there is a k-algebra isomorphism $F_i \cong F$. Therefore, F_i is a radical extension of k. For each j we have $F_j = k(u_{j1}, \ldots, u_{jn})$, where $u_{ji}^{r_i}$ is in $k(u_{j1}, \ldots, u_{j,i-1})$. Therefore

$$K = F_1 F_2 \cdots F_m = k(u_{11}, \ldots, u_{1n}, u_{21}, \ldots, u_{2n}, \ldots, u_{m1}, \ldots, u_{mn})$$

is a radical extension of k. □

THEOREM 5.6.23. *Let k be a field of characteristic zero and assume for each $n > 0$ that $x^n - 1$ splits over k. Let $p(x) \in k[x]$. If $p(x)$ is solvable by radicals over k, then the Galois group of $p(x)$ is a solvable group.*

PROOF. Since $p(x)$ is solvable by radicals, there is a radical tower

$$k = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_n$$

positive integers $r_1, \ldots, r_n$ such that $F_i = F_{i-1}(u_i)$, $u_i^{r_i} \in F_{i-1}$, and $p(x)$ splits over F_n. By Lemma 5.6.22, we can assume F_n is a Galois extension over k. By Kummer Theory (Theorem 5.6.15), F_i is a Galois extension of F_{i-1} and $\mathrm{Aut}_{F_{i-1}} F_i$ is cyclic. By the Fundamental Theorem of Galois Theory (Theorem 5.3.18), F_n is Galois over F_i, $\mathrm{Aut}_{F_i}(F)$ is a normal subgroup of $\mathrm{Aut}_{F_{i-1}}(F)$ and

$$\mathrm{Aut}_{F_{i-1}} F_i \cong \mathrm{Aut}_{F_{i-1}}(F)/\mathrm{Aut}_{F_i}(F).$$

Therefore the series of groups

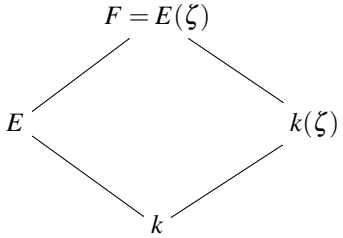
$$\mathrm{Aut}_k F_n \supseteq \mathrm{Aut}_{F_1} F_n \supseteq \cdots \supseteq \mathrm{Aut}_{F_{i-1}} F_n \supseteq \mathrm{Aut}_{F_i} F_n \supseteq \cdots \supseteq \mathrm{Aut}_{F_{n-1}} F_n \supseteq \langle e \rangle$$

is a normal series and at each step the quotient is an abelian group. So the series is a solvable series for $\mathrm{Aut}_k F_n$. Let E be the splitting field for $p(x)$ over k in F_n. Then E is an intermediate field. By Theorem 5.3.15, E is a Galois extension of k. By the Fundamental Theorem of Galois Theory, $\mathrm{Aut}_k E$ is a homomorphic image of $\mathrm{Aut}_k F_n$. By Exercise 2.10.18 , $\mathrm{Aut}_k E$ is solvable. □

Theorem 5.6.24 is a partial converse to Theorem 5.6.23. In characteristic zero, if f is a polynomial with solvable Galois group, then f is solvable by radicals.

THEOREM 5.6.24. *Let k be a field of characteristic zero, $f \in k[x]$ a separable polynomial and E a splitting field for f. If $\mathrm{Aut}_k(E)$ is solvable, then f is solvable by radicals. That is, there exists a radical extension of k that contains E.*

PROOF. Let $n = \dim_k(E)$. Let $F = E(\zeta)$ be a cyclotomic extension of E of order n. That is, ζ is a primitive nth root of unity over k.



By Theorem 5.3.15, E/k is a Galois extension and by hypothesis $\mathrm{Aut}_k(E)$ is a solvable group. By Theorem 5.4.6, $F = E(\zeta)$ is a Galois extension of $k(\zeta)$ and $G = \mathrm{Aut}_{k(\zeta)}(F)$ embeds as a subgroup of $\mathrm{Aut}_k(E)$. By Exercise 2.10.18, G is a solvable group. By Exercise 2.10.20, G has a composition series $G = G_0 \supseteq G_1 \supseteq G_2 \supseteq \cdots \supseteq G_m = \langle e \rangle$ where the factor group G_i/G_{i+1} is cyclic of order $[G_i : G_{i+1}]$, a prime divisor of $|G|$. By Theorem 5.3.18 there is a tower of field extensions $F = F_0 \supseteq F_1 \supseteq F_2 \supseteq \cdots \supseteq F_m = k(\zeta)$ and F_i/F_{i+1} is a cyclic extension, hence a Kummer extension. By Theorem 5.6.15, $F_i = F_{i+1}(v_i)$ is a radical extension. Since $k(\zeta)$ is a radical extension, this proves F/k is a radical extension. □

6.7. Exercises.

EXERCISE 5.6.25. Let k be a field, $n \geq 1$ and $a \in k$. Let $f = x^n - a$ and F/k a splitting field for f. Show that the following are equivalent

(1) Every root of f in F is a simple root.
(2) $F[x]/(f)$ is a direct sum of fields.
(3) $n = 1$ or $na \neq 0$.

EXERCISE 5.6.26. This exercise is a continuation of Exercise 4.2.13. Let R be a UFD with quotient field K. Assume the characteristic of R is not equal to 2. Let $a \in R$ be an element which is not a square in R and $f = x^2 - a \in R[x]$. Let $S = R[x]/(f)$, $L = K[x]/(f)$.

(1) Show that there is a commutative square

$$\begin{array}{ccc} S & \longrightarrow & L \\ \uparrow & & \uparrow \\ R & \longrightarrow & K \end{array}$$

where each arrow is the natural map and each arrow is one-to-one.
(2) Show that L is the quotient field of S.
(3) $\mathrm{Aut}_K L = \langle \sigma \rangle$ is a cyclic group of order two and L/K is a Galois extension.
(4) If $\sigma : L \to L$ is the automorphism of order two, then σ restricts to an R-automorphism of S.
(5) The norm map $N_K^L : L \to K$ restricts to a norm map $N_R^S : S \to R$.

EXERCISE 5.6.27. Let p be a prime number, and F/k an extension of fields which is cyclic of degree p^n. If E is an intermediate field such that $F = E(a)$, and E/k is cyclic of degree p^{n-1}, then $F = k(a)$.

EXERCISE 5.6.28. Let k be a field of positive characteristic p.

(1) The map $a \mapsto a^p - a$ defines a homomorphism of additive groups $\varphi : k \to k$. Prove that a cyclic extension field E/k of degree p exists if and only if the map φ is not onto.
(2) In this exercise, we outline a proof that a cyclic extension field E/k of degree p^{n-1} can be embedded in a cyclic extension field F/k of degree p^n. For the complete classification of cyclic extensions F/k of degree p^n, the interested reader is referred to [**1**]. Assume $n > 1$, E/k is cyclic of degree p^{n-1}, and $\mathrm{Aut}_k(E) = \langle \sigma \rangle$.
   (a) Show that there exists $a, b \in E$ satisfying: $T_k^E(a) = 1$ and $\sigma(b) - b = a^p - a$.
   (b) Show that $x^p - x - a$ is irreducible in $E[x]$.
   (c) Let $F = E[x]/(x^p - x - a)$. Show that F/E is cyclic of degree p and F/k is cyclic of degree p^n.

EXERCISE 5.6.29. Let K be a finite extension field of $\mathbb{Q}$. Prove that K contains only a finite number of roots of unity.

7. Transcendental Field Extensions

For a finite extension of fields K/k we prove that a transcendence base exists and any two transcendence bases have the same number of elements. Therefore, the transcendence degree of the extension is well defined. These notions play important roles in Algebraic Geometry. The field of rational functions K on an algebraic variety V is a finite extension of the ground field k. The transcendence degree of K/k is equal to the dimension of V. In other words, the number of topological degrees of freedom on V is equal to the

number of algebraic degrees of freedom in K. In the Fundamental Theorem on Symmetric Rational Functions we prove that the transcendence degree of the field of symmetric rational functions in n variables over k is equal to n. Moreover, we show that the field of symmetric rational functions is generated by the elementary symmetric polynomials, hence a transcendence base is constructed. In the Fundamental Theorem on Symmetric Polynomials we prove that the ring of symmetric polynomials is generated by the elementary symmetric polynomials. In fact, we show that the ring of symmetric polynomials contains a transcendence base for the field of symmetric rational functions. This is called a globalization result, because rational functions in general have a nonempty pole set, but polynomials do not. There is a version of Emmy Noether's Normalization Lemma (see [**5**, Theorem 8.4.6]) that says under certain sufficient conditions a transcendence base can be constructed globally.

7.1. Transcendence Bases.

DEFINITION 5.7.1. Let F/k be an extension of fields and $\Xi \subseteq F$. We say Ξ is *algebraically dependent* over k if there exist n distinct elements $\xi_1, \ldots, \xi_n$ in Ξ and a nonzero polynomial $f \in k[x_1, \ldots, x_n]$ such that $f(\xi_1, \ldots, \xi_n) = 0$. Otherwise we say Ξ is *algebraically independent*. A *transcendence base* for F/k is a subset $\Xi \subseteq F$ which satisfies

(1) Ξ is algebraically independent over k and
(2) if $\Xi \subseteq Z$ and Z is algebraically independent over k, then $\Xi = Z$.

LEMMA 5.7.2. *Let F/k be an extension of fields and Ξ a subset of F which is algebraically independent over k. For $u \in F - k(\Xi)$, the following are equivalent*

(1) $\Xi \cup \{u\}$ is algebraically independent over k.
(2) u is transcendental over $k(\Xi)$.

PROOF. (2) implies (1): Suppose there exist a polynomial f in $k[x_1, \ldots, x_n]$ and elements $\xi_1, \ldots, \xi_{n-1}$ in Ξ such that $f(\xi_1, \ldots, \xi_{n-1}, u) = 0$. Expand f as a polynomial in x_n with coefficients in $k[x_1, \ldots, x_{n-1}]$, say $f = \sum_j h_j x_n^j$. Then $0 = f(\xi_1, \ldots, \xi_{n-1}, u) = \sum_j h_j(\xi_1, \ldots, \xi_{n-1}) u^j$. But u is transcendental over $k(\Xi)$, so $h_j(\xi_1, \ldots, \xi_{n-1}) = 0$ for each j. But Ξ is algebraically independent, so each polynomial $h_j = 0$. Thus $f = 0$.

(1) implies (2): We prove the contrapositive. Assume u is algebraic over $k(\Xi)$ and $f = \min.\mathrm{poly}_{k(\Xi)}(u) = x^m + h_{m-1}x^{m-1} + \cdots + h_1 x + h_0$. Each h_j is in $k(\Xi)$, so there is a finite subset $\xi_1, \ldots, \xi_n$ of Ξ and polynomials $\alpha_0, \ldots, \alpha_m$, $\beta_0, \ldots, \beta_m$ in $k[x_1, \ldots, x_n]$ such that $h_j = \alpha_j(\xi_1, \ldots, \xi_n)/\beta_j(\xi_1, \ldots, \xi_n)$. Multiply across by the least common multiple, β, of the denominators to get

$$f(x)\beta(\xi_1, \ldots, \xi_n) = \sum_j \gamma_j(\xi_1, \ldots, \xi_n) x^j$$

where $\beta(\xi_1, \ldots, \xi_n) \neq 0$ and each γ_j is in $k[x_1, \ldots, x_n]$. Since $(f\beta)(\xi_1, \ldots, \xi_n, u) = 0$, we are done. □

LEMMA 5.7.3. *Let F/k be an extension of fields and Ξ a subset of F which is algebraically independent over k. The following are equivalent.*

(1) F is algebraic over $k(\Xi)$.
(2) Ξ is a transcendence base for F over k.

PROOF. (1) implies (2): Suppose Z is linearly independent, $Z \supseteq \Xi$, and $z \in Z$. Then z is algebraic over $k(\Xi)$, so by Lemma 5.7.2, $\Xi \cup \{z\}$ is linearly dependent. Therefore, $z \in \Xi$, which implies $Z = \Xi$.

(2) implies (1): We prove the contrapositive. Suppose $u \in F - k(\Xi)$ and u is transcendental over $k(\Xi)$. By Lemma 5.7.2, $\Xi \cup \{u\}$ is algebraically independent, so Ξ is not a transcendence base. □

LEMMA 5.7.4. *Let F be a finitely generated field extension of k. Then the following are true:*

(1) If Ξ is a finite subset of F such that F is algebraic over $k(\Xi)$, then there is a subset of Ξ that is a transcendence base for F/k.
(2) There is a finite transcendence base for F/k.

PROOF. We prove (1). The reader should verify (2). Let Ξ be a finite subset of F such that F is algebraic over $k(\Xi)$. Consider the finite set

$$S = \{Z \subseteq \Xi \mid Z \text{ is algebraically independent over } k\}$$

ordered by set containment. Then S contains a maximal member, call it X. Given $u \in \Xi$, by Lemma 5.7.2, u is algebraic over $k(X)$. By Proposition 5.1.10 (3), $k(\Xi)$ is algebraic over $k(X)$. By Proposition 5.1.10 (4), F is algebraic over $k(X)$. By Lemma 5.7.3, X is a transcendence base. □

THEOREM 5.7.5. *Let F/k be an extension of fields and assume $\Xi = \{\xi_1, \ldots, \xi_n\}$ is a transcendence base for F over k. If Z is another transcendence base for F over k, then Z also has cardinality n.*

PROOF. Step 0: If $n = 0$, then by Exercise 5.7.14, F/k is an algebraic extension. Since Z is algebraically independent over k, we conclude that $Z = \emptyset$. Assume from now on that $n > 0$.

Step 1: There exists $\zeta_1 \in Z$ such that $\zeta_1, \xi_2, \ldots, \xi_n$ is a transcendence base for F/k. First we show that there exists $\zeta \in Z$ such that ζ is transcendental over $K = k(\xi_2, \ldots, \xi_n)$. Assume the contrary. Then F is algebraic over $K(Z)$ and $K(Z)$ is algebraic over K, hence F is algebraic over K. Then ξ_1 is algebraic over K, which contradicts Lemma 5.7.2. Suppose $\zeta_1 \in Z$ and ζ_1 is transcendental over K. By Lemma 5.7.2, $\{\zeta_1, \xi_2, \ldots, \xi_n\}$ is algebraically independent over k. The set $\{\zeta_1, \xi_2, \ldots, \xi_n\} \cup \{\xi_1\}$ is algebraically dependent, so Lemma 5.7.2 says ξ_1 is algebraic over $k(\zeta_1, \xi_2, \ldots, \xi_n)$. In this case, the field $k(\Xi)(\zeta_1) = k(\zeta_1, \xi_2, \ldots, \xi_n)(\xi_1)$ is algebraic over $k(\zeta_1, \xi_2, \ldots, \xi_n)$ and F is algebraic over $k(\Xi)(\zeta_1) = k(\zeta_1, \xi_2, \ldots, \xi_n)(\xi_1)$, hence by Proposition 5.1.10 (4), F is algebraic over $k(\zeta_1, \xi_2, \ldots, \xi_n)$. By Lemma 5.7.3, the set $\zeta_1, \xi_2, \ldots, \xi_n$ is a transcendence base for F/k.

Step 2: Iterate Step 1 $n-1$ more times to get a subset $\{\zeta_1, \ldots, \zeta_n\}$ of Z which is a transcendence base for F/k. By Definition 5.7.1, this implies $Z = \{\zeta_1, \ldots, \zeta_n\}$. □

DEFINITION 5.7.6. Let F/k be an extension of fields such that a finite transcendence base exists. The *transcendence degree* of F/k, denoted $\mathrm{tr.deg}_k(F)$, is the number of elements in any transcendence base of F over k.

THEOREM 5.7.7. *Suppose $k \subseteq F \subseteq K$ is a tower of field extensions. Assume $\Xi = \{\xi_1, \ldots, \xi_n\}$ is a transcendence base for F/k and $Z = \{\zeta_1, \ldots, \zeta_m\}$ is a transcendence base for K/F. Then*

(1) $\Xi \cup Z$ is a transcendence base for K/k, and
(2) $\mathrm{tr.deg}_k(K) = \mathrm{tr.deg}_k(F) + \mathrm{tr.deg}_F(K)$.

PROOF. (2): Follows straight from (1).

(1): The reader should verify that K is algebraic over $k(Z \cup \Xi)(F)$ and $k(Z \cup \Xi)(F)$ is algebraic over $k(Z \cup \Xi)$. Therefore, K is algebraic over $k(Z \cup \Xi)$. Let f be a polynomial

in $k[x_1,\dots,x_n][z_1,\dots,z_m]$ such that $f(\xi_1,\dots,\xi_n,\zeta_1,\dots,\zeta_m)=0$. Since Z is algebraically independent over F, this implies $f(\xi_1,\dots,\xi_n,z_1,\dots,z_m)$ is the zero polynomial in the ring $k(\xi_1,\dots,\xi_n)[z_1,\dots,z_m]$. Therefore, each coefficient of $f(\xi_1,\dots,\xi_n,z_1,\dots,z_m)$ is an algebraic relation over k involving $\xi_1,\dots,\xi_n$. Because $\xi_1,\dots,\xi_n$ are algebraically independent over k, we conclude that $f=0$. This proves $Z\cup\Xi$ is algebraically independent over k. By Lemma 5.7.3 we are done. □

7.2. Symmetric Rational Functions. Let k be a field and $A=k[x_1,\dots,x_n]$ the ring of polynomials over k in the variables $x_1,\dots,x_n$ (see Section 3.6.1). The field of rational functions in $x_1,\dots,x_n$ over k is denoted $K=k(x_1,\dots,x_n)$. Let S_n be the symmetric group on $\{1,2,\dots,n\}$. The group S_n acts on A as a group of k-algebra automorphisms in the following way. Given any permutation $\sigma\in S_n$ and any polynomial $f(x_1,\dots,x_n)\in A$, define $\sigma(f)$ to be the polynomial $f(x_{\sigma(1)},\dots,x_{\sigma(n)})$. Using Theorem 3.6.2 we see that σ defines an automorphism of A that fixes each element of k. By Exercise 3.5.2, the permutation σ induces an automorphism of K and S_n can be viewed as a group of automorphisms of K. Then K is a Galois extension of K^{S_n} with group S_n. The degree of the extension K/K^{S_n} is equal to the order of the group S_n, which is $n!$, by Example 2.1.14. The fixed field K^{S_n} is called the *field of symmetric rational functions in n variables over k*. The subring of A fixed by S_n is denoted A^{S_n}. We call A^{S_n} the *ring of symmetric polynomials in n variables over k*. Let λ be another indeterminate. Consider the polynomial

$$\Lambda=(\lambda-x_1)(\lambda-x_2)\cdots(\lambda-x_n)$$

in $A[\lambda]$. Notice that Λ is symmetric in $x_1,\dots,x_n$. In other words, if we extend the action by S_n on A to an action on the ring $A[\lambda]$, then Λ is fixed by S_n. Therefore, the coefficients of Λ are symmetric polynomials and belong to the ring A^{S_n}. The *elementary symmetric polynomial of degree i in the variables* $x_1,\dots,x_n$, denoted $\sigma_{i,n}$, is the coefficient of λ^{n-i} in the expansion of Λ:

$$\Lambda=\lambda^n-\sigma_{1,n}\lambda^{n-1}+\sigma_{2,n}\lambda^{n-2}-\cdots+(-1)^i\sigma_{i,n}\lambda^{n-i}+\cdots+(-1)^n\sigma_{n,n}.$$

We see that

$$\begin{aligned}
\sigma_{1,n}&=x_1+x_2+\cdots+x_n\\
\sigma_{2,n}&=\sum_{i_1<i_2}x_{i_1}x_{i_2}\\
\sigma_{3,n}&=\sum_{i_1<i_2<i_3}x_{i_1}x_{i_2}x_{i_3}\\
&\vdots\\
\sigma_{n,n}&=x_1x_2\dots x_n
\end{aligned}$$

By Exercise 5.7.18, if $1<i<m\le n$, then the polynomials $\sigma_{i,m}$ satisfy the recurrence relation: $\sigma_{i,m}=\sigma_{i,m-1}+x_m\sigma_{i-1,m-1}$. Therefore, we have the tower of fields: $k(\sigma_{1,n},\dots,\sigma_{n,n})\subseteq k(x_1,\dots,x_n)^{S_n}\subseteq k(x_1,\dots,x_n)$.

THEOREM 5.7.8. *(Fundamental Theorem on Symmetric Rational Functions) Let k be a field and $k(x_1,\dots,x_n)$ the field of rational functions in the variables $x_1,\dots,x_n$ over k. Let S_n be the symmetric group on $\{1,\dots,n\}$ and $k(x_1,\dots,x_n)^{S_n}$ the field of symmetric rational functions in the variables $x_1,\dots,x_n$ over k. Then the following are true.*

(1) $k(x_1,\dots,x_n)$ is a Galois extension of $k(x_1,\dots,x_n)^{S_n}$ with Galois group S_n.
(2) The degree of the extension $k(x_1,\dots,x_n)/k(x_1,\dots,x_n)^{S_n}$ is $n!$.

(3) If $\sigma_{1,n},\dots,\sigma_{n,n}$ are the elementary symmetric polynomials in $x_1,\dots,x_n$, then $k(x_1,\dots,x_n)^{S_n} = k(\sigma_{1,n},\dots,\sigma_{n,n})$.

(4) $k(x_1,\dots,x_n)$ is the splitting field of the polynomial

$$\Lambda = \lambda^n - \sigma_{1,n}\lambda^{n-1} + \sigma_{2,n}\lambda^{n-2} - \cdots + (-1)^i\sigma_{i,n}\lambda^{n-i} + \cdots + (-1)^n\sigma_{n,n}$$

over the field $k(x_1,\dots,x_n)^{S_n} = k(\sigma_{1,n},\dots,\sigma_{n,n})$.

PROOF. Parts (1) and (2) were proved in the paragraph preceding this theorem. By definition, $\Lambda = (\lambda - x_1)(\lambda - x_2)\cdots(\lambda - x_n)$ splits over $k(x_1,\dots,x_n)$ and $k(x_1,\dots,x_n)$ is generated by the roots of Λ. This proves $k(x_1,\dots,x_n)$ is the splitting field for Λ over $k(\sigma_{1,n},\dots,\sigma_{n,n})$, which is (4). By Corollary 5.2.6 and and Corollary 5.2.8, the dimension of $k(x_1,\dots,x_n)$ as a vector space over $k(\sigma_{1,n},\dots,\sigma_{n,n})$ is at most $n!$. Part (2) and Exercise 5.1.23 imply $k(x_1,\dots,x_n)^{S_n} = k(\sigma_{1,n},\dots,\sigma_{n,n})$, which proves (3). □

COROLLARY 5.7.9. *Let k be a field and $k[x_1,\dots,x_n]$ the ring of polynomials in the variables $x_1,\dots,x_n$ over k. If $\sigma_{1,n},\dots,\sigma_{n,n}$ are the elementary symmetric polynomials in $x_1,\dots,x_n$, then the k-algebra homomorphism $k[t_1,\dots,t_n] \to k[\sigma_{1,n},\dots,\sigma_{n,n}]$ defined by $t_i \mapsto \sigma_{i,n}$ is an isomorphism.*

PROOF. By Exercise 5.7.16, $K = k(x_1,\dots,x_n)$ has transcendence degree n over k. By Theorem 5.7.8, K is algebraic over $k(s_{1,n},\dots,s_{n,n})$. By Lemma 5.7.4 and Theorem 5.7.5, $\{s_{1,n},\dots,s_{n,n}\}$ is a transcendence base for K over k. Therefore, the k-algebra homomorphism $k[t_1,\dots,t_n] \to k[s_{1,n},\dots,s_{n,n}]$ defined by $t_i \mapsto \sigma_{i,n}$ is a k-algebra isomorphism. □

COROLLARY 5.7.10. *If G is a finite group, then there exists a Galois field extension with Galois group isomorphic to G.*

PROOF. Let $[G:1] = n$. By Cayley's Theorem, Theorem 2.4.4, we can identify G with a subgroup of S_n. By Theorem 5.7.8 and Theorem 5.3.18, $k(x_1,\dots,x_n)$ is a Galois extension of $k(x_1,\dots,x_n)^G$ with Galois group G. □

7.3. The General Polynomial of Degree n is not solvable by Radicals. Let k be a field of characteristic zero and assume $x^d - 1$ splits over k, for each $d > 1$. Let $t_0, t_1, \dots, t_{n-1}$ be indeterminates, and $K = k(t_0, t_1, \dots, t_{n-1})$ the field of rational functions over k. The general polynomial of degree n over the field k is

$$p(x) = x^n - t_{n-1}x^{n-1} + \dots (-1)^{n-1}t_1x + (-1)^n t_0$$

which is an element of the ring $K[x]$.

COROLLARY 5.7.11. *If $n \geq 5$, the general polynomial of degree n is not solvable by radicals.*

PROOF. Let $\sigma_1,\dots,\sigma_n$ be the elementary symmetric polynomials in the n variables $x_1,\dots,x_n$. By Theorem 5.7.8, $K = k(x_1,\dots,x_n)$ is the splitting field of the polynomial

$$\begin{aligned}\Lambda &= (\lambda - x_1)(\lambda - x_2)\cdots(\lambda - x_n)\\ &= \lambda^n - \sigma_{1,n}\lambda^{n-1} + \cdots + (-1)^{n-1}\sigma_{n-1,n}\lambda + (-1)^n\sigma_{n,n}.\end{aligned}$$

over $k(\sigma_{1,n},\dots,\sigma_{n,n})$. By Corollary 5.7.9, the field $k(\sigma_{1,n},\dots,\sigma_{n,n})$ is isomorphic to the field of rational functions $k(t_0,t_1,\dots,t_{n-1})$ in n variables over k. Therefore, Λ is a general polynomial of degree n over k. The Galois group of K over $k(\sigma_{1,n},\dots,\sigma_{n,n})$ is S_n, the symmetric group on n letters. By Corollary 2.10.14, S_n is not solvable. By Theorem 5.6.23, Λ is not solvable by radicals, □

7.4. Symmetric Polynomials. Theorem 5.7.8 (3) says that every symmetric rational function is a rational function in the elementary symmetric polynomials. In Theorem 5.7.12, which is due to Gauss, we improve this result by proving that every symmetric polynomial is a polynomial in the elementary symmetric polynomials.

THEOREM 5.7.12. *(Fundamental Theorem on Symmetric Polynomials) Let k be a field and $k[x_1,\dots,x_n]$ the ring of polynomials in the variables $x_1,\dots,x_n$ over k. Let S_n be the symmetric group on $\{1,\dots,n\}$ and $k[x_1,\dots,x_n]^{S_n}$ the ring of symmetric polynomials in the variables $x_1,\dots,x_n$ over k. If $\sigma_{1,n},\dots,\sigma_{n,n}$ are the elementary symmetric polynomials in $x_1,\dots,x_n$, then the following are true.*

(1) If f is a nonzero symmetric polynomial, then there exists a polynomial $g \in k[t_1,\dots,t_n]$ such that $f = g(\sigma_{1,n},\dots,\sigma_{n,n})$.
(2) $k[x_1,\dots,x_n]^{S_n} = k[\sigma_{1,n},\dots,\sigma_{n,n}]$.
(3) The polynomial g in (1) is unique.

The proof of the theorem will utilize the following lemma.

LEMMA 5.7.13. *In the context of Theorem 5.7.12, let f be a nonzero symmetric polynomial in $k[x_1,\dots,x_n]^{S_n}$. If the leading term of f (see Lemma 3.6.16) is $M = rx_1^{e_1}\cdots x_n^{e_n}$, then $e_1 \geq e_2 \geq \cdots \geq e_n$.*

PROOF. For sake of contradiction assume $1 \leq i < j \leq n$ and $e_i < e_j$. Apply the transposition $\tau = (i,j)$ to f. Since $\tau f = f$, we know that f has the monomial

$$\tau M = rx_1^{e_1}\cdots x_{i-1}^{e_{i-1}} x_j^{e_i} x_{i+1}^{e_{i+1}} \cdots\cdots x_{j-1}^{e_{j-1}} x_i^{e_j} x_{j+1}^{e_{j+1}} \cdots x_n^{e_n} = rx_1^{e_1}\cdots x_i^{e_j}\cdots x_j^{e_i}\cdots x_n^{e_n}.$$

Thus in the monomial τM, the exponents of x_i and x_j are swapped. But

$$M = rx_1^{e_1}\cdots x_i^{e_i}\cdots x_j^{e_j}\cdots x_n^{e_n} < rx_1^{e_1}\cdots x_i^{e_j}\cdots x_j^{e_i}\cdots x_n^{e_n} = \tau M.$$

This is a contradiction, since M is the leading term of f. □

PROOF OF THEOREM 5.7.12. (1) and (2): Let $f \in k[x_1,\dots,x_n]^{S_n}$ be a nonzero symmetric polynomial and assume the leading term of f is $r_1 x_1^{e_1}\cdots x_n^{e_n}$. By Lemma 5.7.13, $e_1 \geq e_2 \geq \cdots \geq e_n$. Set $d_1 = e_1 - e_2$, $d_2 = e_2 - e_3$, …, $d_{n-1} = e_{n-1} - e_n$, and $d_n = e_n$. By Exercise 5.7.20, the leading term of $s_{1,n}^{d_1} s_{2,n}^{d_2}\cdots s_{n,n}^{d_n}$ is equal to

$$x_1^{d_1+d_2+\cdots+d_n} x_2^{d_2+\cdots+d_n}\cdots x_n^{e_n} = x_1^{e_1} x_2^{e_2}\cdots x_n^{e_n}.$$

Let $g_1 = r_1 s_{1,n}^{d_1} s_{2,n}^{d_2}\cdots s_{n,n}^{d_n}$. Then $g_1 \in k[s_{1,n},\dots,s_{n,n}]$ and $f_1 = f - g_1$ is a symmetric polynomial in $k[x_1,\dots,x_n]^{S_n}$. The leading terms of f and g_1 are equal, so if f_1 is nonzero, the leading term of f_1 is less than the leading term of f in the lexicographical order (see Section 3.6.1). If f_1 is nonzero, then we repeat the above steps to get $g_2 \in k[s_{1,n},\dots,s_{n,n}]$ with the same leading term as f_1. Hence $f_2 = f_1 - g_2$ is either zero, or has a leading term less than the leading term of f_1. Iterating, we get a sequence of symmetric polynomials $f, f_1, f_2, \dots$ such that the leading terms form a strictly decreasing sequence. By Lemma 3.6.16 (3), after a finite number of iterations we have $f_m = 0$. This shows that $f = g_1 + g_2 + \cdots + g_m$ is in $k[s_{1,n},\dots,s_{n,n}]$, proving (1) and (2).

(3): This follows from Corollary 5.7.9, because the map induced by sending t_i to $\sigma_{i,n}$ is a k-algebra isomorphism $k[t_1,\dots,t_n] \cong k[s_{1,n},\dots,s_{n,n}]$. □

7.5. Exercises.

EXERCISE 5.7.14. If F/k is an extension of fields, show that $\emptyset$ is a transcendence base if and only if F/k is an algebraic extension.

EXERCISE 5.7.15. If F/k is an extension of fields, and $\Xi \subseteq F$ is algebraically independent over k, show that there exists a transcendence base Z such that $Z \supseteq \Xi$.

EXERCISE 5.7.16. Let k is a field, and $x_1,\ldots,x_n$ a set of indeterminates. Show that $\operatorname{tr.deg}_k k(x_1,\ldots,x_n) = n$ and $\{x_1,\ldots,x_n\}$ is a transcendence base for $k(x_1,\ldots,x_n)$ over k.

EXERCISE 5.7.17. If F is a finitely generated extension field of the field k, show that $\operatorname{tr.deg}_k(F)$ is equal to the least integer n such that there exist $\xi_1,\ldots,\xi_n$ in F and F is algebraic over $k(\xi_1,\ldots,\xi_n)$.

EXERCISE 5.7.18. Let $x_1,\ldots,x_n$ be a set of indeterminates. If $1 \leq i \leq m \leq n$, let $\sigma_{i,m}$ be the elementary symmetric polynomial of degree i in the variables $x_1,\ldots,x_m$. Prove the following recursive formula:

$$\sigma_{i,m} = \begin{cases} x_1 + x_2 + \cdots + x_m & \text{if } i = 1, \\ x_1 x_2 \ldots x_m & \text{if } i = m, \\ \sigma_{i,m-1} + x_m \sigma_{i-1,m-1} & \text{if } 1 < i < m \leq n. \end{cases}$$

EXERCISE 5.7.19. Let S_n be the symmetric group on $\{1,2,\ldots,n\}$ and S_{n-1} the symmetric group on $\{1,2,\ldots,n-1\}$. We view S_{n-1} as a subgroup of S_n. Let k be a field. Prove that if $f(x_1,\ldots,x_n) \in k[x_1,\ldots,x_n]^{S_n}$, then $f(x_1,\ldots,x_{n-1},0) \in k[x_1,\ldots,x_{n-1}]^{S_{n-1}}$. Show that there exists a commutative diagram

$$\begin{array}{ccc} A_n = k[x_1,\ldots,x_n] & \xrightarrow{\alpha} & A_{n-1} = k[x_1,\ldots,x_{n-1}] \\ {\scriptstyle a}\uparrow\subseteq & & {\scriptstyle b}\uparrow\subseteq \\ A_n^{S_n} & \xrightarrow{\beta} & A_{n-1}^{S_{n-1}} \\ {\scriptstyle c}\uparrow\subseteq & & {\scriptstyle d}\uparrow= \\ k[\sigma_{1,n},\ldots,\sigma_{n,n}] & \xrightarrow{\gamma} & k[\sigma_{1,n-1},\ldots,\sigma_{n-1,n-1}] \end{array}$$

of commutative rings satisfying the following:

(1) The maps a,b,c,d are homomorphisms defined by set inclusion.
(2) The epimorphism α is defined by $x_n \mapsto 0$.
(3) The homomorphism β is the restriction of α to $A_n^{S_n}$.
(4) The epimorphism γ is the restriction of α to $k[\sigma_{1,n},\ldots,\sigma_{n,n}]$.

EXERCISE 5.7.20. Let $e_i \geq 0$ for each i. In the context of Theorem 5.7.12, show that the leading term of $s_{1,m}^{e_1} s_{2,m}^{e_2} \cdots s_{m,m}^{e_m}$ is equal to $x_1^{e_1+e_2+\cdots+e_m} x_2^{e_2+\cdots+e_m} \cdots x_m^{e_m}$.

EXERCISE 5.7.21. Follow the steps below to show that the map γ in Exercise 5.7.19 has a section.

(1) Show that there is a k-algebra homomorphism

$$\varepsilon : k[\sigma_{1,n-1},\ldots,\sigma_{n-1,n-1}] \to k[\sigma_{1,n},\ldots,\sigma_{n,n}]$$

defined by $\sigma_{i,n-1} \mapsto \sigma_{i,n}$.
(2) Show that $\gamma\varepsilon$ is the identity map on $k[\sigma_{1,n-1},\ldots,\sigma_{n-1,n-1}]$.

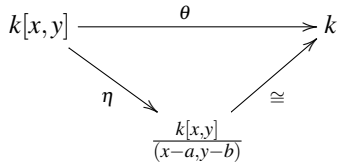
EXERCISE 5.7.22. Let F/k be an extension of fields. Apply Zorn's Lemma, Proposition 1.3.3, to prove: If Ξ is a subset of F such that F is algebraic over $k(\Xi)$, then Ξ contains a subset which is a transcendence base for F over k.

8. Applications to Algebraic Curves

If k is a field, then the affine plane over k is the cartesian product $k^2 = k \times k$. If x and y are indeterminates and $f(x,y) \in k[x,y]$, then

$$Z(f) = \{(a,b) \in k^2 \mid f(a,b) = 0\}$$

is the set of solutions of the equation $f(x,y) = 0$ in k^2. We call $Z(f)$ an affine algebraic curve. This terminology agrees with that of Section 5.1.2 where we discussed lines and circles in the affine plane. The commutative ring $R = k[x,y]/(f)$ is known as the affine coordinate ring of the curve $Z(f)$. There is a correspondence between points (a,b) on the cuve $Z(f)$ and maximal ideals in R. For instance, given $(a,b) \in Z(f)$, consider the ideal $(x-a, y-b)$ in $k[x,y]$. Applying Exercise 3.6.34 twice, once for $x-a$ and once for $y-b$, we see that $k[x,y]/(x-a,y-b) \cong k$. Hence $(x-a,y-b)$ is a maximal ideal. Applying Theorem 3.6.2 for x and y, there is a k-algebra homomorphism $\theta : k[x,y] \to k$ defined by $x \mapsto a$ and $y \mapsto b$. Clearly θ is onto and the maximal ideal $(x-a,y-b)$ is contained in $\ker\theta$. Hence $\ker\theta = (x-a,y-b)$. By Theorem 3.2.15, there is a commutative diagram



of k-algebras where η is the natural map. Since $\theta(f) = f(a,b) = 0$, $f \in \ker\theta = (x-a, y-b)$. If we set $\mathfrak{m} = (x-a,y-b)$ in R, then by Corollary 3.2.16, $R/\mathfrak{m} = k[x,y]/(x-a,y-b) = k$ and $\mathfrak{m}$ is a maximal ideal of R. The correspondence between points of $Z(f)$ and maximal ideals of R is not onto. That is, if M is a maximal ideal of R, then R/M is in general an extension field of k, hence M does not have the form $(x-a,y-b)$ and does not correspond to a point on $Z(f)$. In this case, the maximal ideal M is called an R/M-rational point of the curve $Z(f)$. In the example of Section 5.8.1 below, we show that when k is not algebraically closed, the affine coordinate ring R of the circle $Z(x^2+y^2-1)$ has maximal ideals such that the residue field R/M is strictly greater than k (see Proposition 5.8.1 (4)). Although we do not prove it here, since R/M is a finitely generated k-algebra, by the Hilbert Basis Theorem, R/M is finitely generated and algebraic over k (see [**4**, Proposition 10.2.4] or [**2**, Proposition 7.9], for example). Therefore, if k is algebraically closed, then every point of $Z(f)$ is k-rational.

8.1. A Nonsingular Affine Conic. If k is a field, then the unit circle C in the plane k^2 is the set of solutions of the equation $x^2+y^2-1=0$. That is,

$$C = \{(x,y) \in k^2 \mid x^2+y^2-1=0\}.$$

This terminology agrees with that of Section 5.1.2. In this section we investigate the commutative ring $R = k[x,y]/(x^2+y^2-1)$ which is known as the affine coordinate ring of the unit circle C.

First we establish notation that will be in effect throughout this section. Let k be a field such that x^2+1 is irreducible over k. In particular, this implies that the characteristic of the base field k is not 2 (Exercise 3.2.30). Let $k[x]$ be the polynomial ring in one variable

over k. Then $k[x]$ is a UFD (Example 3.4.12) and $x-1$ is a prime in $k[x]$. Let $k(x)$ be the field of rational functions, the quotient field of $k[x]$. Consider the polynomial x^2+y^2-1 in $k[x][y]$. By Eisenstein's Criterion (Theorem 3.7.6) with prime $p=x-1$, $y^2+(x^2-1)$ is irreducible in $k[x][y]$. By Theorem 3.7.5, the polynomial ring $k[x][y]$ is a UFD. Therefore,

$$R=\frac{k[x,y]}{(x^2+y^2-1)}$$

is an integral domain, by Corollary 3.4.14. The ring R is known as the affine coordinate ring of the unit circle C in the affine plane k^2. By Gauss' Lemma (Theorem 3.7.4), x^2+y^2-1 is irreducible in $k(x)[y]$ and

$$F=\frac{k(x)[y]}{(y^2+x^2-1)}$$

is a field. By Exercise 5.3.32, F is a Galois extension of $k(x)$. The Galois group $\mathrm{Aut}_{k(x)}F$ is cyclic of order 2, and generated by the automorphism τ defined by $y\mapsto -y$. Let $K=k(i)$ be the splitting field for x^2+1 over k. The Galois group of K/k is the cyclic group $\langle\sigma\rangle$, where $\sigma(i)=-i$. In the following, cosets in the factor rings R and F are written without brackets or any extra adornment.

PROPOSITION 5.8.1. *In the above context, the following properties hold for R and F:*

1. *F is the quotient field of R.*
2. *As a $k[x]$-module, R is free of rank two with basis $1,y$.*
3. *There is a norm map $N^R_{k[x]}: R\to k[x]$ defined by $a+by\mapsto (a+by)(a-by)=a^2-b^2y^2=a^2-b^2(1-x^2)$.*
4. *In general, R contains K-rational points.*

PROOF. (1): The diagram of ring homomorphisms

(8.1)
$$\begin{array}{ccc} R=\frac{k[x,y]}{(x^2+y^2-1)} & \xrightarrow{\phi} & F=\frac{k(x)[y]}{(x^2+y^2-1)} \\ \uparrow{\scriptstyle \eta} & & \uparrow{\scriptstyle \eta} \\ k[x,y] & \xrightarrow{\alpha} & k(x)[y] \\ \uparrow & & \uparrow \\ k[x] & \longrightarrow & k(x) \end{array}$$

commutes. The vertical maps are the natural maps. The horizontal map α exists by Theorem 3.6.2 applied to $k[x]\to k(x)$. Since $\eta\alpha(x^2+y^2-1)=0$, ϕ exists by Theorem 3.2.15. Using Gauss' Lemma (Theorem 3.7.4), we see that the kernel of $\eta\alpha$ is the principal ideal (x^2+y^2-1). Therefore, ϕ is one-to-one. By Exercise 3.5.2, we can view the quotient field of R as a subfield of F. In this context, we show that F is equal to the quotient field of R. By Lemma 4.4.3, a $k(x)$-basis for F is $\{1,y\}$. Since $y\in R$ we know y is in the quotient field of R. The quotient field of $k[x]$ is $k(x)$, hence $k(x)$ is in the quotient field of R. A typical element of F is of the form $f(x)+g(x)y$, where $f(x)$ and $g(x)$ are in $k(x)$. Hence a typical element of F is in the quotient field of R.

(2): By Lemma 4.4.3, a $k(x)$-basis for F is $\{1,y\}$. Therefore, $\{1,y\}$ is linearly independent over $k[x]$. Since the image of $\eta\alpha$ is generated by polynomials over k in the variables x and y, $\{1,y\}$ is a generating set for the image of ϕ as a $k[x]$-module. In Diagram (8.1), ϕ is one-to-one. So $\{1,y\}$ is a generating set for R as an A-module.

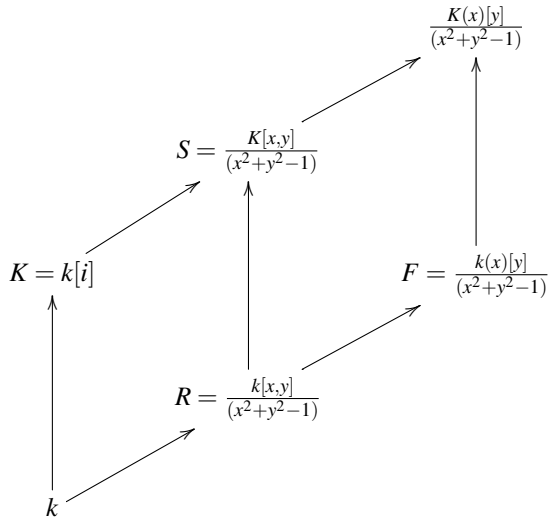
(3): The norm map $N^F_{k(x)}: F\to k(x)$ restricts to a norm map $R\to k[x]$.

(4): To see that for general k, R has K-rational points, suppose for instance that 2 is a square in k. Then $R/(y-\sqrt{2}) \cong k[x]/(x^2+1) \cong K$. Therefore, the principal ideal $(y-\sqrt{2})$ is a maximal ideal of R with residue field K. □

We retain the notation from Proposition 5.8.1. The affine coordinate ring of the unit circle in the plane K^2 is $S = K[x,y]/(x^2+y^2-1)$. Identifying K with $k[z]/(z^2+1)$, we see that

$$\begin{aligned} S &= \frac{k[x,y,z]}{(x^2+y^2-1, z^2+1)} \\ &= \frac{R[z]}{(z^2+1)} \\ &= R[i]. \end{aligned}$$

The diagram

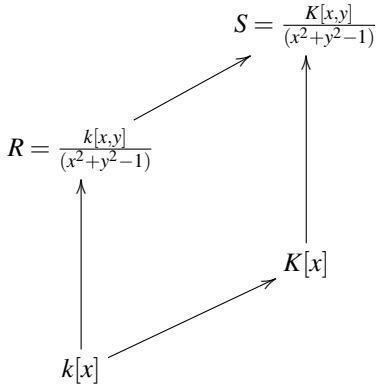


commutes. By Proposition 5.8.1, the quotient field of S is

$$\frac{K[x,y]}{(x^2+y^2-1)} = \frac{k[x,y,z]}{(x^2+y^2-1, z^2+1)} = F[i].$$

The field extension $F[i]/F$ is Galois with group $\langle\sigma\rangle$ where $\sigma(i) = -i$. Notice that σ restricts to an R-algebra automorphism of S and the norm $N_R^S : S^* \to R^*$ is a homomorphism of groups. We can also view R as the ramified quadratic extension of $k[x]$ defined by adjoining the square root of $1-x^2$. Likewise, S is the ramified quadratic extension of $K[x]$

defined by adjoining the square root of $1-x^2$.



Geometrically, the extension $R/k[x]$ or $S/K[x]$ corresponds to the projection of the circle C onto the x-axis.

PROPOSITION 5.8.2. *In the above context, the following are true.*

(1) S is a UFD.
(2) $S^* = K^* \times \langle x+iy \rangle$.
(3) $R^* = k^*$.
(4) R is not a UFD. In fact, x, $y+1$, $y-1$ are irreducible in R and $x^2 = (y+1)(y-1)$.

PROOF. (1) and (2): To show that S is a UFD, by Exercise 3.6.26, it suffices to show that S is isomorphic to the ring of Laurent polynomials $K[u,u^{-1}]$. In S we have $x^2+y^2-1 = (x+iy)(x-iy)-1$. Define K-algebra homomorphisms

$$\frac{K[u,v]}{(uv-1)} \xrightarrow{\phi} \frac{K[x,y]}{(x^2+y^2-1)} \xrightarrow{\theta} \frac{K[u,v]}{(uv-1)}$$

by $\phi(u) = x+iy$, $\phi(v) = x-iy$, $\theta(x) = \frac{u+v}{2}$ and $\theta(y) = \frac{u-v}{2i}$. One can check that ϕ and θ are well defined K-algebra homomorphisms. Since ϕ and θ are inverses of each other, they are isomorphisms. There is an isomorphism of K-algebras

$$\frac{K[u,v]}{(uv-1)} \xrightarrow{\cong} K[u,u^{-1}]$$

induced by $v \mapsto u^{-1}$. By Exercise 3.6.26, $K[u,u^{-1}]$ is a UFD and the group of units is equal to the internal direct product $K^* \times \langle u \rangle$. Using the isomorphism θ, this proves S is a UFD and

$$S^* = K^* \times \langle x+iy \rangle.$$

Notice that the inverse of $x+iy$ is $x-iy$. This proves (1) and (2).

(3): We have the homomorphism of groups $N_R^S : S^* \to R^*$ and if $a \in R^*$, then $N_R^S(a) = a^2$. Since $N_R^S(x+iy) = (x+iy)(x-iy) = x^2+y^2 = 1$, we see that $R^* = (K^*)^{\langle \sigma \rangle} = k^*$.

(4): To prove that x is irreducible in R, we use the norm map $R \to k[x]$ of Proposition 5.8.1 (3). Look at the norm of x from R to $k[x]$:

$$N_{k[x]}^R(x) = x^2.$$

For sake of contradiction, assume x has a nontrivial factorization $x = \alpha\beta$ in R. Since $R^* = k^*$, this means the norm of α is equal to cx for some $c \in k^*$. Suppose $\alpha = a+by$ for

some $a,b \in k[x]$. Then the equation $N^R_{k[x]}(a+by) = cx$ becomes

$$a^2 - b^2(1-x^2) = cx.$$

Substitute $x = 1$ and $x = -1$ to get $c = a(1)^2$ and $-c = a(-1)^2$. Hence

$$-1 = a(1)^2 a(-1)^{-2}$$

which contradicts our assumption that -1 is not a square in k. Therefore, x is irreducible in R. Since

$$N^R_{k[x]}(1+y) = N^R_{k[x]}(1-y) = (1+y)(1-y) = x^2$$

the same argument proves that $y+1$ and $y-1$ are irreducible in R. This proves R is not a UFD, since the identity

$$x^2 = (1-y)(1+y)$$

holds in R. □

PROPOSITION 5.8.3. *In the above context, the ideal $\mathfrak{m} = (x, y-1)$ of R has the following properties:*

(1) $\mathfrak{m}$ is a maximal ideal.
(2) $\mathfrak{m}$ is not a principal ideal.
(3) $\mathfrak{m}^2$ is equal to the principal ideal $(y-1)$.

PROOF. (1): Since $R/\mathfrak{m} = k[x,y]/(x,y-1) = k$ is a field, $\mathfrak{m}$ is a maximal ideal.

(2): Assume $\mathfrak{m} = (z)$ is a principal ideal. Then z divides x. Since x is irreducible, this implies z and x are associates. But $R/(x) = k[y]/(y^2-1)$ is not a field. Therefore, $\mathfrak{m} \neq (x)$, a contradiction.

(3): Notice that $\mathfrak{m}^2 = (x, y-1)^2$ is generated by the three elements $x^2 = 1-y^2$, $x(y-1)$, and $(y-1)^2$, all of which are in the principal ideal $(y-1)$. Conversely, since $x^2+y^2 = 1$ in R,

$$\begin{aligned} x^2 + (y-1)^2 &= x^2 + y^2 - 2y + 1 \\ &= 2(1-y) \end{aligned}$$

which shows $y-1$ is in $\mathfrak{m}^2$. This proves $\mathfrak{m}^2 = (y-1)$ is a principal ideal in R. □

PROPOSITION 5.8.4. *In the above context, the ideal $\mathfrak{m} = (x, y-1)$ of R has the following properties:*

(1) $\mathfrak{m}$ is a projective R-module.
(2) $\mathfrak{m}$ is not a free R-module.

PROOF. (1): From Proposition 5.8.1, we can view R as a subring of F. An arbitrary element $m \in \mathfrak{m}$ can be written in the form $m = ax + b(y-1)$, for some $a,b \in R$. From

$$\begin{aligned} \left(\frac{y+1}{x}\right) m &= \frac{y+1}{x}(ax + b(y-1)) \\ &= \frac{ax(y+1) + b(y^2-1)}{x} \\ &= \frac{ax(y+1) - bx^2}{x} \\ &= a(y+1) - bx \end{aligned}$$

we see that $\left(\frac{y+1}{x}\right) m \in R$. Define

$$\begin{aligned} \mathfrak{m} &\xrightarrow{\phi} R^2 \\ m &\mapsto \left(\frac{y+1}{2x}m, \frac{-m}{2}\right) \end{aligned}$$

and

$$\begin{aligned} R^2 &\xrightarrow{\pi} \mathfrak{m} \\ (a,b) &\mapsto ax + b(y-1). \end{aligned}$$

The reader should verify that ϕ and π are R-module homomorphisms. For each $m \in \mathfrak{m}$ we have

$$\begin{aligned} \pi\phi(m) &= \pi\left(\frac{y+1}{2x}m, \frac{-m}{2}\right) \\ &= \left(\frac{y+1}{2x}\right) mx - \frac{m}{2}(y-1) \\ &= \frac{y+1}{2}m - \frac{y-1}{2}m \\ &= m. \end{aligned}$$

Therefore, $\pi\phi = 1_{\mathfrak{m}}$. Hence ϕ is one-to-one and π is onto. By Proposition 4.1.21, $\phi(M)$ is an R-module direct summand of R^2. By Proposition 4.1.29, M is a finitely generated projective R-module.

(2): For sake of contradiction, assume $\mathfrak{m}$ is a free R-module of rank r. By Exercise 4.2.16, $\mathfrak{m}/\mathfrak{m}^2$ is a vector space of dimension r over the field $R/\mathfrak{m}$. By Proposition 5.8.3 (3), $\mathfrak{m}/\mathfrak{m}^2$ is generated by x. Therefore, $r = 1$. This implies $\mathfrak{m}$ is a principal ideal, contradicting Proposition 5.8.3 (2). □

8.2. A Nonsingular Affine Elliptic Curve. This short section is devoted to an example of an algebraic curve that is nonsingular and nonrational. Assume that the characteristic of k, the base field, is not 2. Let $A = k[x]$ be the polynomial ring in one variable over k. Then A is a UFD (Example 3.4.12) and x is a prime in A. Let $K = k(x)$ be the quotient field of A. Consider the polynomial $y^2 - x(x^2-1)$ in $A[y]$. By Eisenstein's Criterion (Theorem 3.7.6) with prime $p = x$, $y^2 - x(x^2-1)$ is irreducible in $A[y]$. By Gauss' Lemma (Theorem 3.7.4), $y^2 - x(x^2-1)$ is irreducible in $K[y]$ and $F = K[y]/(y^2 - x(x^2-1))$ is a field. By Exercise 5.3.32, F/K is a Galois extension, $\mathrm{Aut}_K(F) = \langle\sigma\rangle$ has order 2, and σ is defined by $y \mapsto -y$. The norm map is $N_K^F : F \to K$.

In the following, cosets in the factor ring F are written without brackets or any extra adornment. By Theorem 3.7.5, the polynomial ring $A[y] = k[x,y]$ is a UFD. Therefore, $R = k[x,y]/(y^2 - x(x^2-1))$ is an integral domain, by Corollary 3.4.14. The diagram of

ring homomorphisms

$$\begin{array}{ccc} A = k[x] & \longrightarrow & K = k(x) \\ \downarrow & & \downarrow \\ A[y] & \xrightarrow{\alpha} & K[y] \\ \eta \downarrow & & \downarrow \eta \\ R = A[y]/(y^2 - x(x^2-1)) & \xrightarrow{\phi} & F = K[y]/(y^2 - x(x^2-1)) \end{array} \tag{8.2}$$

commutes. The vertical maps are the natural maps. The horizontal map α exists by Theorem 3.6.2 applied to $A \to K$. Since $\eta\alpha(y^2 - x(x^2-1)) = 0$, ϕ exists by Theorem 3.2.15. Using Gauss' Lemma (Theorem 3.7.4), we see that the kernel of $\eta\alpha$ is the principal ideal $(y^2 - x(x^2-1))$. Therefore, ϕ is one-to-one.

PROPOSITION 5.8.5. *In the above context, the following are true.*

(1) The quotient field of R is F.
(2) As an A-module, R is free of rank 2. The set $\{1, y\}$ is a free basis. The image of ϕ is $\{p(x) + q(x)y \mid$ where $p(x)$ and $q(x)$ are in $A = k[x]\}$.
(3) The homomorphism $A \to R$ defined by sending x to its image in R is one-to-one.
(4) The automorphism $\sigma \in \mathrm{Aut}_K(F)$ defined by $y \mapsto -y$ restricts to an automorphism $\sigma : R \to R$.
(5) For any $a \in R$, define the norm of a to be $N(a) = a\sigma(a)$. Then $N(1) = 1$, $N : R \to A$, and N is multiplicative.
(6) The map on groups of units $k^* \to R^*$ is an isomorphism. That is, the units of R are precisely the units of k.
(7) x and y are irreducible elements of R.
(8) R is not a unique factorization domain.
(9) R is not a principal ideal domain.

PROOF. (1): By Exercise 3.5.2, we can view the quotient field of R as a subfield of F. In this context, we show that F is equal to the quotient field of R. By Lemma 4.4.3, a $k(x)$-basis for F is $\{1, y\}$. Since $y \in R$ we know y is in the quotient field of R. The quotient field of $k[x]$ is $k(x)$, hence $k(x)$ is in the quotient field of R. A typical element of F is of the form $f(x) + g(x)y$, where $f(x)$ and $g(x)$ are in $k(x)$. Hence a typical element of F is in the quotient field of R.

(2): By Lemma 4.4.3, a K-basis for F is $\{1, y\}$. Therefore, $\{1, y\}$ is linearly independent over A. Since the image of $\eta\alpha$ is generated by polynomials over k in the variables x and y, $\{1, y\}$ is a generating set for the image of ϕ as an A-module. As mentioned in the paragraph that precedes the proposition, ϕ is one-to-one. So $\{1, y\}$ is a generating set for R as an A-module.

(3): The composite map $A \to K \to F$ is one-to-one and factors through R.

(4): Using Theorem 3.6.2, we see that the map $\sigma : A[y] \to A[y]$ defined by $y \mapsto -y$ is an automorphism and maps the principal ideal $(y^2 - x(x^2-1))$ onto itself.

$$\begin{array}{ccc} A[y] & \xrightarrow{\sigma} & A[y] \\ \eta \downarrow & & \downarrow \eta \\ R & \longrightarrow & R \end{array} \tag{8.3}$$

The kernel of $\eta\sigma$ is the principal ideal $(y^2 - x(x^2-1))$. Hence $\sigma : R \to R$ is an automorphism.

(5): Let $a \in R$. By (2), a has a unique representation in the form $a = f + gy$, for polynomials f and g in $A = k[x]$. Then $N(a) = a\sigma(a) = f^2 - g^2y^2 = f^2 - g^2x(x^2-1)$ is in the image of $A \to R$. Notice that $N : R \to A$ is the restriction of $N_K^F : F \to K$, hence $N(1) = 1$ and $N(ab) = N(a)N(b)$ by Section 5.5.

(6): The map $k \to R$ is one-to-one because k is a field. We show that $k^* \to R^*$ is onto. Let $a, b \in R$ and assume $ab = 1$. Then $N(a)N(b) = 1$ in A. But $A^* = k^*$. This proves $N(a) \in k$. By (2), a has a unique representation in the form $a = f + gy$, for polynomials f and g in $A = k[x]$. Then $N(a) = f^2 - g^2x(x^2-1) = u$ for some $u \in k^*$. Then $(f(0))^2 = u$. If $g \neq 0$, then the leading term of f^2 which is even is equal to the leading term of $g^2x(x^2-1)$, which is odd, a contradiction. Therefore, $g = 0$ and $a = f = f(0)$ is in k.

(7): If x is not irreducible, then there is a nontrivial factorization $x = ab$. By (5), we have the factorization $N(x) = x^2 = N(a)N(b)$ in $A = k[x]$. Therefore, $N(a) = x$ up to associates. By (2), a has a representation in the form $a = f + gy$, for polynomials f and g in $A = k[x]$. Then up to associates, $N(a) = f^2 - g^2x(x^2-1) = x$. Then $f^2 = g^2x(x^2-1) + x$ which is impossible because the degree of the left hand is even and that of the right hand side is odd. This proves x is not in the image of the norm map $N : R \to A$, hence x is irreducible in R.

If y is not irreducible in R, then there is a nontrivial factorization $y = ab$. By (5), we have the factorization $N(y) = x(x^2-1) = N(a)N(b)$ in $A = k[x]$. Therefore, up to associates, one of $N(a)$ or $N(b)$ is in $\{x, x+1, x-1\}$. The same proof from above shows that $x+1$ and $x-1$ are not in the image of $N : R \to A$. Therefore, y is irreducible in R.

(8): In R we have the identity $y^2 = x(x^2-1)$. By the proof of (7), $N(x) = x^2$ and $N(y) = x(x^2-1)$. Therefore, x and y are not associates of each other. So unique factorization does not exist.

(9): Consider the ideal $\mathfrak{m} = (x, y)$. Then $R/\mathfrak{m} = k[x,y]/(x,y) = k$ is a field, hence $\mathfrak{m}$ is a maximal ideal. If $\mathfrak{m} = (a)$ is principal, then $a \mid x$ and $a \mid y$. Since x and y are irreducible, by Lemma 3.4.5, this implies x and y are associates of each other, a contradiction to (8). □

8.3. Exercises.

EXERCISE 5.8.6. Let k be a field. Assume the characteristic of k is not 2 or 3 and that k contains a primitive sixth root of unity denoted ζ_6.

(1) Show that $k(x)$ is a cyclic Galois extension of $k(x^6)$ of degree 6 (in other words, a Kummer extension). Let $G = \langle\sigma\rangle$ be the Galois group. Determine the lattice of subfields and lattice of subgroups guaranteed by the Fundamental Theorem of Galois Theory.
(2) Show that G acts on $k[x]$ and the fixed subring is $k[x^6]$. Determine the lattice of fixed subrings of $k[x]$ corresponding to the subgroups of G.
(3) As in Exercise 3.6.17, let $R = k[x^2, x^3]$. Show that the quotient field of R is $k(x)$. We say that R is *birational to* $k[x]$. Determine the subgroup of G that fixes R point-wise (that is, the stabilizer of R in G).
(4) True or False?
   (a) $k[x]$ is a free $k[x^2]$-module.
   (b) $k[x]$ is a free $k[x^2, x^3]$-module.
   (c) $k[x^2, x^3]$ is a free $k[x^2]$-module. (Hint: $k[x^2]$ is a PID, in fact it is a euclidean domain. Section 4.3 applies.)

EXERCISE 5.8.7. Let k be a field. In Algebraic Geometry, the ring $k[x^2,x^3]$ of Exercise 3.6.17 corresponds to a cuspidal cubic curve and is not a UFD. The ring $k[x^2,x+x^3]$ corresponds to a nodal cubic curve.

(1) Show that the quotient field of $k[x^2,x+x^3]$ is $k(x)$. In other words, $k[x^2,x+x^3]$ and $k[x]$ are birational.
(2) Prove that $k[x^2,x+x^3]$ is not a UFD.

EXERCISE 5.8.8. In the context of Proposition 5.8.5, consider the maximal ideal $\mathfrak{m} = (x,y)$. Show that $\mathfrak{m}^2$ is principal.

Acronyms

ACC Ascending Chain Condition
DCC Descending Chain Condition
GCD Greatest Common Divisor
PID Principal Ideal Domain
UFD Unique Factorization Domain

Bibliography

[1] A. A. Albert, *Cyclic fields of degree p^n over F of characteristic p*, Bull. Amer. Math. Soc. **40** (1934), no. 8, 625–631. MR 1562919

[2] M. F. Atiyah and I. G. Macdonald, *Introduction to commutative algebra*, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1969. MR 0242802 (39 #4129)

[3] Allan Clark, *Elements of abstract algebra*, Dover Publications, Inc., New York, 1971, unabridged and Corrected republication of the work first published by Wadsworth Publishing Company, Belmont, California, in 1971.

[4] Timothy J. Ford, *Abstract algebra*, pre-preprint. Available at: `http://math.fau.edu/ford/preprints/Algebra_Book/Algebra_Book.pdf`.

[5] ———, *Separable algebras*, Graduate Studies in Mathematics, vol. 183, American Mathematical Society, Providence, RI, 2017. MR 3618889

[6] Évariste Galois, *Écrits et mémoires mathématiques*, Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics], Éditions Jacques Gabay, Paris, 1997, Édition critique intégrale des manuscrits et publications. [Integral critical edition of the manuscripts and publications], With a preface by Jean Dieudonné, Edited, with notes and commentary by Robert Bourgne and Jean-Pierre Azra, Reprint of the second (1976) edition. MR 1452597

[7] I. N. Herstein, *Topics in algebra*, second ed., Xerox College Publishing, Lexington, Mass., 1975. MR 0356988 (50 #9456)

[8] Thomas W. Hungerford, *Algebra*, Graduate Texts in Mathematics, vol. 73, Springer-Verlag, New York, 1980, Reprint of the 1974 original. MR 600654 (82a:00006)

[9] John L. Kelley, *General topology*, Springer-Verlag, New York-Berlin, 1975, Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.], Graduate Texts in Mathematics, No. 27. MR 0370454 (51 #6681)

[10] James H. McKay, *Another proof of Cauchy's group theorem*, Amer. Math. Monthly **66** (1959), 119. MR 98777

[11] Eugen Netto, *Ueber die Irreductibilität ganzzahliger ganzer Functionen*, Math. Ann. **48** (1896), no. 1-2, 81–88. MR 1510925

[12] David Singmaster and D. M. Bloom, *Problems and Solutions: Solutions of Elementary Problems: E1648*, Amer. Math. Monthly **71** (1964), no. 8, 918–920. MR 1532917

[13] Michael Spivak, *Calculus*, fourth ed., Publish or Perish, Inc., PMB 377, 1302 Waugh Drive, Houston, Texas 77019, 2008.

[14] The Sage Development Team, *Sagemath, the Sage Mathematics Software System (Version 8.8)*, The Sage Development Team, 2019-06-26, `http://www.sagemath.org`.

[15] A. R. Wadsworth, *Problems in abstract algebra*, Student Mathematical Library, vol. 82, American Mathematical Society, Providence, RI, 2017. MR 3643210

[16] Helmut Wielandt, *Ein Beweis für die Existenz der Sylowgruppen*, Arch. Math. (Basel) **10** (1959), 401–402. MR 147529

[17] Max Zorn, *A remark on method in transfinite algebra*, Bull. Amer. Math. Soc. **41** (1935), no. 10, 667–670. MR 1563165

Glossary of Notations

$\langle X \mid Y \rangle$	group defined by generators X and relations Y, 58
(X)	ideal generated by X, 91
(X)	submodule generated by X, 127
1_X	identity map on X, 10
2^X	power set of X, 9
$[G:1]$	order of the group G, 27
$[G:H]$	index of the subgroup H in the group G, 34
$\mathrm{annih}_R(M)$	annihilator of M in R, 126
$\mathrm{Aut}(G)$	group of automorphisms of G, 39
$\mathrm{Aut}_R(A)$	automorphism group of an R-algebra A, 143
$\bigcap_{i\in I} X_i$	intersection of a family of sets, 9
$\bigcup_{i\in I} X_i$	union of a family of sets, 9
$\binom{n}{k}$	binomial coefficient, 12
$\mathrm{char}(R)$	characteristic of R, 91
$\dim_D(V)$	dimension of the D-vector space V, 136
$\ell(G)$	length of a composition series of G, 83
$\emptyset$	empty set, 9
$\gcd(a_1,\ldots,a_n)$	greatest common divisor of $\{a_1,\ldots,a_n\}$, 15
$\mathrm{GL}_n(F)$	general linear group of n-by-n matrices over F, 30
$\mathrm{Hom}(A,A)$	endomorphism ring of an abelian group, 86
$\mathrm{im}\,(f)$	image of a homomorphism f, 90
$\ker(f)$	kernel of a homomorphism f, 90
$\langle X \rangle$	subgroup generated by X, 33
$\lceil x \rceil$	ceiling of x, 14
$\mathrm{lcm}(a,b)$	least common multiple, 17
$\lfloor x \rfloor$	floor of x, 14
$\lvert a \rvert$	order of the element a, 27
$\lvert X \rvert$	cardinality of the set X, 11
$\mathrm{Map}(X)$	set of all functions from X to X, 13
$\mathbb{C}$	complex numbers, 9
$\mathbb{N}$	natural numbers, 9
$\mathbb{N}_n$	$\{1,2,\ldots,n\}$, 12
$\mathbb{Q}$	rational numbers, 9
$\mathbb{Q}/\mathbb{Z}$	rational numbers modulo the integers, 42
$\mathbb{Q}/\mathbb{Z}(p)$	p-torsion subgroup of $\mathbb{Q}/\mathbb{Z}$, 142
$\mathbb{R}$	real numbers, 9
$\mathbb{Z}$	integers, 9
$\mathbb{Z}/(m)$	integers modulo m, 17
$\mathrm{min.poly}_k(\alpha)$	minimal polynomial of α over k, 144

μ	the group of all roots of unity in $\mathbb{C}$, 42
μ_n	the group of nth roots of unity, 42
$\mathrm{Perm}(X)$	set of all permutations of X, 25
$\mathrm{PGL}_n(F)$	projective general linear group, 46
$\phi(n)$	Euler ϕ-function, 18
$\prod_{i\in I} X_i$	product of a family of sets, 22
ψ^*	the dual of $\psi \in \mathrm{Hom}_R(M,N)$, 148
$\mathrm{Rad}_R(0)$	nil radical of R, 97
$\mathrm{Rank}(M)$	rank of the module M, 131
$\mathrm{sign}(\sigma)$	sign of a permutation, 62
$\mathrm{SL}_n(F)$	special linear group, 46
$\mathrm{trace}(\alpha)$	trace of a matrix, 170
$\mathrm{trace}(\phi)$	trace of a homomorphism, 170
$\mathrm{tr.deg}_k(F)$	transcendence degree of F/k, 216
$\mathrm{Units}(R)$ or R^*	group of units in the ring R, 86
$a+I$	left coset of I containing a, 93
$a \mid b$	a divides b, 15
A^n	powers of an ideal, 98
A_n	alternating group on n letters, 62
D_n	dihedral group of order $2n$, 29
e_{ij}	elementary matrix, 97
$F(X)$	free group on the set X, 58
f'	formal derivative of the polynomial f, 115
$G \cong G'$	G is isomorphic to G', 27
G/H	set of all left cosets of H modulo G, 34
G'	commutator subgroup, 48
G^o	opposite group, 31
$H \rtimes K$	semidirect product of H and K, 53
$H\backslash G$	set of all right cosets of H modulo G, 34
$I:J$	ideal quotient, 97
$k(x)$	field of rational functions over k in the variable x, 113
$M(\phi,X,Y)$	matrix of ϕ with respect to the bases X, Y, 146
$M(\pi)$	submodule of M annihilated by powers of π, 140
M/S	factor module of M modulo S, 128
M^*	$\mathrm{Hom}_R(M,R)$, the dual module, 148
$M_1 \oplus M_2 \oplus \cdots \oplus M_n$	direct sum of modules, 130
$M_n(R)$	ring of n-by-n matrices over R, 86
$M_{nm}(R)$	set of all n-by-m matrices over R, 146
$N \trianglelefteq G$	N is a normal subgroup of G, 39
$N_G(S)$	normalizer of S in G, 52
$o(G)$	order of the group G, 27
Q_8	quaternion eight group, 29
$R(G)$	group ring, 86
R/I	residue class ring, 93
R^o	opposite ring of R, 87
$R^{(n)}$	free R-module of rank n, 131
$S_1+S_2+\cdots+S_n$	sum of submodules, 130
S_n	symmetric group on n letters, 12

Index